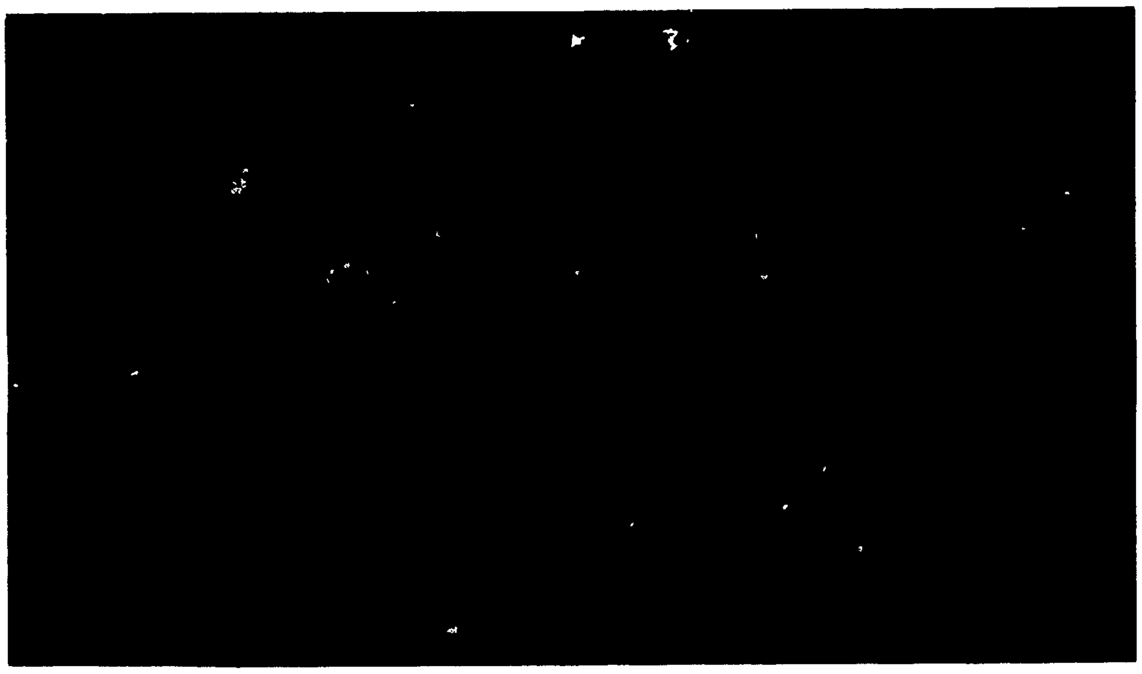
DOCUMENT RESUME

ED 035 956	24	CG 005 020

AUTHOR	REMSTAD, ROBERT C.; AND OTHERS
TITLE	OPTIMIZING THE RESPONSE TO A CONCEPT ATTAINMENT TEST THROUGH SEQUENTIAL CLASSROOM EXPERIMENTATION. TECHNICAL REPORT NO. 85.
INSTITUTION	WISCONSIN UNIV., MADISON. RESEARCH AND DEVELOPMENT CENTER FOR COGNITIVE LEARNING.
SPONS AGENCY	OFFICE OF EDUCATION (DHEW), WASHINGTON, D.C. BUREAU OF RESEARCH.
REPORT NO	TR-85
BUREAU NO	BR-5-0216
PUB DATE	APR 69
CONTRACT	OEC-5-10-154
NOTE	147P.
EDRS PRICE	EDRS PRICE MF-$0.75 HC-$7.45
DESCRIPTORS	*CONCEPT FORMATION, *ELEMENTARY SCHOOL STUDENTS, EXPERIMENTAL SCHOOLS, EXPERIMENTS, LABORATORY TECHNIQUES, *LEARNING, RESEARCH, *RESPONSE MODE, TEACHING TECHNIQUES, *VERBAL STIMULI

ABSTRACT

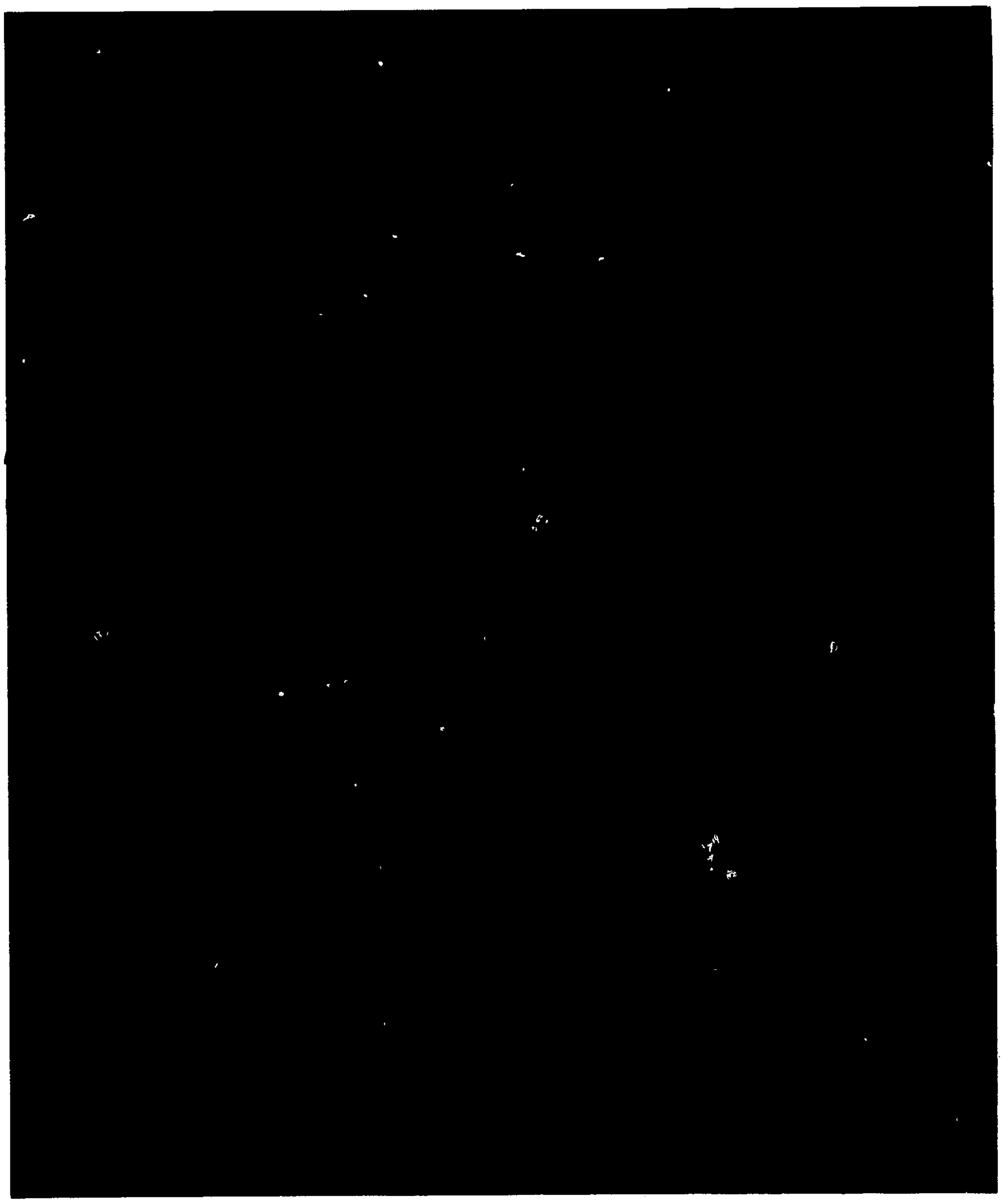
THIS STUDY TESTED THE POSSIBILITY OF USING THE OPTIMIZATION TECHNIQUES OF THE RESPONSE SURFACE MODEL IN MAXIMIZING THE RESPONSE TO A CONCEPT LEARNING TASK. A SECONDARY PURPOSE WAS THE INVESTIGATION OF THE EXTENT TO WHICH FINDINGS FROM CONCEPT LEARNING EXPERIMENTS IN LABORATORY SETTINGS WOULD GENERALIZE TO THE LEARNING OF TYPES OF CONCEPTS TYPICALLY TAUGHT IN SCHOOL. THE CONCEPT LEARNING TASKS USED CONSISTED OF THE LEARNING OF A SERIES OF PLANE GEOMETRY CONCEPTS. A SEQUENCE OF 16 EXPERIMENTS WERE RUN USING FIFTH GRADE CHILDREN FROM 16 SMALL TOWN ELEMENTARY SCHOOLS. WHILE THE SEQUENCE OF EXPERIMENTS DID NOT PROVE LONG ENOUGH TO PRODUCE A MAXIMUM RESPONSE, THE FOLLOWING RESULTS WERE OBTAINED. THE VARIABLES OF VERBAL CUE, TIME PER SLIDE (VISUAL), AND RATIO OF POSITIVE TO NEGATIVE INSTANCES PRODUCED THE GREATEST INCREMENTS IN RESPONSE. THE DESIGN DID MOVE IN A DIRECTION OF A MAXIMAL RESPONSE, ALTHOUGH NOT AS RAPIDLY AS EXPECTED. DESIRABLE FEATURES OF THIS EXPERIMENTAL APPROACH, SUCH AS THE ABILITY TO EXPAND OR CONTRACT THE NUMBER OF VARIABLES, WERE DEMONSTRATED. (AUTHOR)



No. 85

OPTIMIZING THE RESPONSE TO A CONCEPT ATTAINMENT TEST THROUGH SEQUENTIAL CLASSROOM EXPERIMENTATION

Report from the Technical Section



U.S. Office of Education Center No. C-03 Contract OE 5-10-154

Technical Report No. 85

OPTIMIZING THE RESPONSE TO A CONCEPT ATTAINMENT TEST THROUGH SEQUENTIAL CLASSROOM EXPERIMENTATION

Report from the Technical Section

By Robert C. Remstad

Frank B. Baker, Professor of Educational Psychology,
Chairman of the Examining Committee

Mary R. Quilling, Technical Section Director

Wisconsin Research and Development
Center for Cognitive Learning
The University of Wisconsin
Madison, Wisconsin

April 1969

This Technical Report is a doctoral dissertation reporting research supported by the Wisconsin Research and Development Center for Cognitive Learning. Since it has been approved by a University Examining Committee, it has not been reviewed by the Center. It is published by the Center as a record of some of the Center's activities and as a service to the student. The bound original is in The University of Wisconsin Memorial Library.

The research reported herein was performed pursuant to a contract with the United States Office of Education, Department of Health, Education and Welfare, under the provisions of the Cooperative Research Program. The opinions expressed in this publication do not necessarily reflect the position or policy of the Office of Education and no official endorsement by the Office of Education should be inferred.

Center No. C-03 / Contract OE 5-10-154

NATIONAL EVALUATION COMMITTEE

Samuel Brownell
Professor of Urban Education
Graduate School
Yale University

Henry Chauncey
President
Educational Testing Service

Elizabeth Koontz
President
National Education Association

Patrick Suppes
Professor
Department of Mathematics
Stanford University

Launor F. Carter
Senior Vice President on Technology and Development
System Development Corporation

Martin Deutsch
Director, Institute for Developmental Studies
New York Medical College

Roderick McPhee
President
Punahou School, Honolulu

***Benton J. Underwood**
Professor
Department of Psychology
Northwestern University

Francis S. Chase
Professor
Department of Education
University of Chicago

Jack Edling
Director, Teaching Research Division
Oregon State System of Higher Education

G. Wesley Sowards
Director, Elementary Education
Florida State University

UNIVERSITY POLICY REVIEW BOARD

Leonard Berkowitz
Chairman
Department of Psychology

John Guy Fowlkes
Director
Wisconsin Improvement Program

Herbert J. Klausmeier
Director, R & D Center
Professor of Educational Psychology

M. Crawford Young
Associate Dean
The Graduate School

Archie A. Buchmiller
Deputy State Superintendent
Department of Public Instruction

Robert E. Grinder
Chairman
Department of Educational Psychology

Donald J. McCarty
Dean
School of Education

***James W. Cleary**
Vice Chancellor for Academic Affairs

H. Clifton Hutchins
Chairman
Department of Curriculum and Instruction

Ira Sharkansky
Associate Professor of Political Science

Leon D. Epstein
Dean
College of Letters and Science

Clauston Jenkins
Assistant Director
Coordinating Committee for Higher Education

Henry C. Weinlick
Executive Secretary
Wisconsin Education Association

EXECUTIVE COMMITTEE

Edgar F. Borgatta
Brittingham Professor of Sociology

Russell J. Hosler
Professor of Curriculum and Instruction and of Business

Wayne Otto
Professor of Curriculum and Instruction (Reading)

Richard L. Venezky
Assistant Professor of English and of Computer Sciences

Max R. Goodson
Professor of Educational Policy Studies

***Herbert J. Klausmeier**
Director, R & D Center
Professor of Educational Psychology

Robert G. Petzold
Associate Dean of the School of Education
Professor of Curriculum and Instruction and of Music

FACULTY OF PRINCIPAL INVESTIGATORS

Ronald R. Allen
Associate Professor of Speech and of Curriculum and Instruction

Gary A. Davis
Associate Professor of Educational Psychology

Max R. Goodson
Professor of Educational Policy Studies

Richard G. Morrow
Assistant Professor of Educational Administration

Vernon L. Allen
Associate Professor of Psychology
(On leave 1968-69)

M. Vere DeVault
Professor of Curriculum and Instruction (Mathematics)

Warren O. Hagstrom
Professor of Sociology

Wayne Otto
Professor of Curriculum and Instruction (Reading)

Nathan S. Blount
Associate Professor of English and of Curriculum and Instruction

Frank H. Farley
Assistant Professor of Educational Psychology

John G. Harvey
Associate Professor of Mathematics and Curriculum and Instruction

Milton O. Pella
Professor of Curriculum and Instruction (Science)

Robert C. Calfee
Associate Professor of Psychology

John Guy Fowlkes (Advisor)
Professor of Educational Administration
Director of the Wisconsin Improvement Program

Herbert J. Klausmeier
Director, R & D Center
Professor of Educational Psychology

Thomas A. Romberg
Assistant Professor of Mathematics and of Curriculum and Instruction

Robert E. Davidson
Assistant Professor of Educational Psychology

Lester S. Golub
Lecturer in Curriculum and Instruction and in English

Burton W. Kreitlow
Professor of Educational Policy Studies and of Agricultural and Extension Education

Richard L. Venezky
Assistant Professor of English and of Computer Sciences

MANAGEMENT COUNCIL

***Herbert J. Klausmeier**
Director, R & D Center
Acting Director, Program 1

Thomas A. Romberg
Director
Programs 2 and 3

James E. Walter
Director
Dissemination Section

Dan G. Woolpert
Director
Operations and Business

Mary R. Quilling
Director
Technical Section

* COMMITTEE CHAIRMAN

STATEMENT OF FOCUS

The Wisconsin Research and Development Center for Cognitive Learning focuses on contributing to a better understanding of cognitive learning by children and youth and to the improvement of related educational practices. The strategy for research and development is comprehensive. It includes basic research to generate new knowledge about the conditions and processes of learning and about the processes of instruction, and the subsequent development of research-based instructional materials, many of which are designed for use by teachers and others for use by students. These materials are tested and refined in school settings. Throughout these operations behavioral scientists, curriculum experts, academic scholars, and school people interact, insuring that the results of Center activities are based soundly on knowledge of subject matter and cognitive learning and that they are applied to the improvement of educational practice.

The Technical Section, a support activity, functions to identify and invent research and development strategies, to assist in the conduct of research and development pro-

grams, and to train graduate students and other research personnel in research and development strategies. This Technical Report describes a contribution of the Technical Section staff toward the invention of research and development strategies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

This study tested the possibility of using the optimization techniques of the response surface model (which visualizes the totality of responses to all possible combinations of k independent variables as a surface or maniford in a (k+1)-dimensional space) in maximizing the response to a concept learning task. This experimental approach is unique in that a sequence of experiments are performed with the combinations of levels of independent variables comprising the later experiments being determined by the results of the earlier experiments.

A secondary purpose was the investigation of the extent to which findings from concept learning experiments in laboratory settings would generalize to the learning of types of concepts typically taught in schools. The concept learning tasks used consisted of the learning of a series of plane geometry concepts, quadrilateral, isosceles triangle, trapezoid and rectangle, presented by series of slides and synchronized tape recorded verbal cues. The response, mean level of performance on a recognition test given 24 hours after the slide presentation, was seen as a function of these variables: 1) amount of redundant information, 2) mode of presentation of successive instances, 3) ratio of numbers of positive to negative instances, 4) order of positive and negative

instances, 5) accompanying verbal cues, 6) time slides were available to subjects, 7) time between successive slides, and 8) relative complexity of the concept.

A sequence of sixteen experiments were run using fifth grade children from sixteen small town elementary schools. While the sequence of experiments did not prove long enough to produce a maximum response, the following results were obtained.

The variables of verbal cue, time per slide, and ratio of positive to negative instances produced the greatest increments in response.

The design did move in the direction of a maximal response, although not as rapidly as expected. Desirable features of this experimental approach, such as, the ability to expand or contract the number of variables, were demonstrated.

CHAPTER I

NATURE OF THE PROBLEM

Human behavior, even in its simplest and least complicated psychomotor forms, tends to be involved. The cognitive aspects of human behavior are even more complex. While recognizing that a basic purpose of science is to cut away the superfluous, to abstract and analyze into essentials and, in general, to aim at simple explanations of phenomena being studied, it is nonetheless true that researchers often attack problems within an overly simple theoretical and experimental framework solely because the tools do not exist for conveniently handling the problem in all its complexities. Thus, the scientific study of psychology and education has repeatedly resorted to simple models including but a few of the many admittedly interrelated variables which might be determinants of the behavior under study. The ability to study human behavior in appropriately complex ways is a necessity for the increasing understanding and prediction of that behavior.

Statisticians working in the physical sciences and in industry have been concerned for a number of years with analogous problems in attempting to optimize production outputs which are seen as complex functions of

many input variables and in estimating the nature of these complex functional relationships. Some of the more interesting and promising ideas have appeared in what has come to be known as response surface methodology. While the physical scientist has the advantages of working with materials that are highly homogeneous and subject to experimental manipulation free from ethical restraints, and with variables that can be tightly controlled and accurately measured on continuous, interval or ratio scales, this differs from the situation of the behavioral scientist more in degree than in kind.

Less than twenty-five years ago Hotelling wrote:

> The possibilities of improvement of physical and chemical investigations based on the theory of statistical inference have scarcely begun to be explored. *

While today this statement is no longer true with regard to the fields of which Hotelling was speaking, it can still be applied to much of current educational and psychological experimentation.

As Baker (1967) has pointed out, at a time when adequate financing is readily available for large scale, sustained investigation of broad problem areas, educational experimentation is still taking place largely

---

*Hotelling, Harold. Some improvements in weighting and other experimental techniques. Ann. Math. Stat., 1944, 15, 297.

within the framework of a theory of experimental design developed to solve the difficulties of relatively small, stand-alone, agricultural experiments. While these techniques are still valid, they, in general, are difficult to apply to the complex, multifaceted type of research problems currently being funded. New techniques are needed with the capability to handle the sequential investigation of complex and often ill-defined problems. Among the capabilities required are those of integrating knowledge from earlier sub-experiments in modifying later ones, of being able to introduce new variables into the system and rejecting those which seem unimportant, and of adapting to changing criterion measures. One promising area of investigation is the above mentioned response surface approach.

Statement of the Problem

This dissertation is the result of a simultaneous interest in, and concern for, two quite different problem areas, one methodological and the other substantive. The methodological interest is with the topic of response surface methodology and the problem of applying this powerful technique to situations where the response of interest is some type of human behavior. The substantive concern is with the area of concept attainment or learning. The analogy between industrial production and

psychological production seemed a particularly apt one and led to this attempt to use the methodology developed in one in experimentation with the other.

Box and Wilson (1951) published a basic theoretical paper concerned with the optimization of responses in the chemical industry, upon which Box and others (Box, 1954; Box and Hunter, 1957) have since elaborated and enlarged. As they stated the problem one is interested in a response which is supposed dependent upon a number of quantitative factors or variables, each capable of exact measurement and control. The response is subject to error. These k independent variables plus the response measure define a (k+1)-dimensional space. Each combination of variables with the attendant expected response determines a point in this space. The totality of all such points forms a surface over the k-dimensional subspace spanned by the independent variables. This surface is the response surface.

Within the entire multidimensional space one may identify a region, delimited by practical considerations, which Box terms the experimental region. The problem of exploring this experimental region has two aspects. The first is that of determining the approximate subregion in which the responses are optimal. This is accomplished by designing a sequence of experiments

which permit one to move from any arbitrary starting point toward points of maximum or minimum response. The procedure used is generally some variation of a steepest ascent procedure first suggested by Box and Wilson (1951). The second aspect is that of examining this subregion of optimal response with the purpose of estimating the functional relationship between the independent variables and the response within this neighborhood. This is accomplished by saturating the subregion with a sufficient number of experimental points to allow estimation of the function by a polynomial of minimally sufficient degree. Knowledge of this functional relationship results in a more accurate determination of the optimal combination and permits one to set production levels based on the relative costs of maintaining the optimal levels of the various factors.

Box (1957) has carried this idea even further, suggesting that full-scale industrial production processes have continuing experimentation built into them. Small systematic variations in the independent or input variables can be programmed into regular production runs according to a specified experimental design. These variations, while within limits imposed by quality control, are yet large enough to generate detectable effects. These effects can then be analyzed and used to

provide the basis for computing new experimental combinations aimed at optimizing production.

Educators and educational psychologists have analogous problems in attempting to maximize the amount of learning which takes place in classroom and laboratory. There can be little doubt that the problem of maximizing learning is, within certain practical limitations, of major concern to educators, if not the major concern. Obviously the correspondences to industry are not one to one, and any insistence on such correspondences would be tenuous at best. The learning process is perhaps far more complicated than those found in industrial production, functional relationships are less well understood, variables are often discrete rather than continuous, measurement scales are frequently nominal or ordinal in nature, and the raw material is far less tractable to experimental manipulation. Nevertheless, both in education and psychology, new techniques for dealing experimentally with multifactor or multivariate problems are needed and the analogy between industrial production and educational production is close enough to suggest that the response surface model may have value. Furthermore, the increasing control of the stimulus inputs to the learning situation which are certain to result as the full impact of the cybernetic revolution

reaches education will lead to even closer analogies between the two types of production. The major difficulty is in formulating approximations which will permit the use of techniques developed for continuous data with discrete variables measured on an ordinal or quasi-interval metric.

One area of educational and psychological investigation in which a large number of important variables have been identified and to which the response surface model has relevance is in that of concept attainment or concept learning. In recent years the psychological literature has had a plethora of experimental studies concerned with various aspects of how humans attain, acquire, develop, form, identify or learn concepts. While some semantic difficulties exist over this variety of verbs and over the lack of an authoritative definition of the word, concept, it is probably accurate to state that most of these studies pertain to varied aspects of a single, highly complex, cognitive process. Further, there can be no question that a basic goal of education and educational psychology is the maximization of the results of this process.

Klausmeier et al. (1965) in identifying and categorizing variables which appeared to be determinants of, or related to, adequate concept learning behavior,

compiled a list of more than one hundred variables. Many of these, although discrete in nature, are quantitative rather than qualitative variables and thus meet the basic requirement of the response surface model. A number of these, related in that all are aspects of the presentation of stimulus materials, seemed particularly appropriate for this purpose and constitute the independent variables studied in the empirical part of this paper.

A further aspect of the problems being investigated stems from difficulties with the nature of much of the recent concept learning experimentation as it related to education. An examination of the scores of experimental studies devoted to various aspects of concept learning leads to the conclusion that psychologists have compiled a considerable amount of information which should be relevant to the learning of the types of concepts taught in schools. However, closer examination raises the question of whether or not that which the psychologist in his experimentation has labeled a concept bears an adequate enough resemblance to the type of concept taught in the classroom to be of value to educators.

It appears to this writer that there are decisive differences between the two definitions of concept. A

major difficulty in identifying the exact points of difference is the lack of exact definition of what is meant by the word concept, either by the psychologist or by the educator. Further complicating the semantic problem is the fact that concepts are viewed as something to be learned, formed, attained, identified, evolved, acquired or developed, and that these processes may or may not be synonomous, depending upon the researcher. Identical experimental tasks are often used as experimental analogs of psychological experiences which the individual researchers see as different.

Psychological experimentation on the problems of how one attains a concept has relied, to a considerable extent, upon variations of a single type of stimulus material. Reference here is made to the familiar sets of cards representing all possible combinations of two or more figures, kind of shapes, colors, textures, borders, etc. One finds reference to the WCST (Wisconsin Card Sorting Task), the NYU Card Sorting Test, and many other unique variations which appear only once in the literature. Some experimenters have presented the cards in sequence while others have presented the entire array simultaneously. However, with all this surface variability all the tasks still retain basic underlying similarities indicating a widely accepted common idea

of what the experimental psychologist means by concept. That common idea is that a concept is an arbitrary rule for dichotomizing a universe of stimuli into those stimuli which belong and those which do not.

On the following major points these experimental tasks, and the definition implicit in them, provide an inadequate experimental analog for the behavior of the classroom learning of a concept. Each of these sets of stimulus materials is, first of all, restricted in size. A small, finite number of dimensions or attributes are all that are allowed and, typically, only two to four distinguishable values on each of these dimensions. The result is a complete set of instances consisting of all possible combinations of one value from each dimension which usually numbers from sixteen to one hundred twenty-eight unique combinations. Smaller sets tend to be trivial for the college populations from which most subjects have been drawn and larger sets tend to be so difficult to work with that subjects cannot solve the problems presented in reasonable time limits.

A second characteristic of these stimulus materials is that the materials may be readily identified by the subject in words which are part of his existing vocabulary. For example, square, diamond, circle, red, blue, one, two, large, small, etc., are all words which

are readily available to most grade school youngsters, not to mention the college students who represent the majority of subjects in concept attainment or concept learning experiments.

Frequently, the subject is presented the stimulus material in an arbitrary, but highly systematic, order which emphasizes the various dimensions on which the set of materials are organized. This order also may operate to reduce the memory load of a subject attempting to master the task and may even suggest a strategy or procedure for attack upon the task.

The usual task set for the subject in a concept learning experiment is the identification of an arbitrary rule which experimenter has used to classify the stimulus materials into sets of exemplars and non-exemplars of the "concept." This may be an explicit statement of the rule or may be inferred from the subject's demonstrated ability to perform as if he knew the rule, as in a card sorting task. Thus, the subjects in these experiments are doing nothing more than attempting to infer a rule for dichotomizing the stimulus materials into those which belong to some arbitrary subset and those which do not. Since the dimensionality of the system is finite, it is always possible to unambiguously come up with the exact rule for categorizing.

These tasks differ from those faced by a youngster trying to learn a concept in school in at least these three basic aspects. First, the youngster must learn a new word or a new use for a familiar word along with the concept. A vocabulary problem exists which is not a part of most laboratory tasks. Second, neither the exemplars of the concept nor the dimensions to be considered can be enumerated completely, let alone observed in their entirety, so that an unambiguous or complete attainment of a concept is impossible. Finally, the acquisition of a new concept usually calls for the student to categorize his environment in, what to him, is some novel fashion whereas categorizing geometric figures as red or black, large or small, square or circle can hardly be called novel for any population older than six or seven.

Thus, there exists the need for a more realistic experimental analog to concept learning behavior as experienced by children in schools. A task is needed incorporating the learning of a label along with the concept; a task which has built into it an indeterminateness which prevents the learning of a concept in any complete way and which forces the subject to organize relevant aspects of his environment in somewhat novel ways.

Related to these difficulties with the typical concept learning task is a similar kind of problem with the response measures used. When a teacher says that a child has attained, formed or learned a new concept, he is indicating that the child has added something of some permanence to his cognitive repertoire. The response measure of interest to that teacher is some form of retention measure rather than one of immediate learning. Yet in the great majority of concept learning studies the response measure used is some concommitant of immediate learning.

Purpose of the Study

A situation exists where researchers in education have need of methodological tools capable of handling more complex and less well-defined formulations of problems. At the same time, techniques of considerable promise exist in other disciplines. Response surface methodology seems to be one of these techniques of promise. Hill and Hunter (1966), in a very complete survey of response surface literature, cite 49 empirical applications of the model. While these 49 applications are primarily in chemistry, production problems with machine tools, truck tires, coastal bermuda grass, and pie crusts are included. Only one of the 49 references to applications of the model pertains to the study of human

or animal behavior and that article (Meyer, 1963) is essentially a statement that the method should be useful in the study of behavior rather than an application.

The situation in the concept learning literature is comparable. Much experimental work has been done in this area using experimental tasks which may be highly inappropriate analogs of the concept learning which goes on in schools. Little or no empirical work exists which attempts to bridge the gap between the laboratory experiment and the learning activities of the classrooms.

Given this context, this dissertation has as its purpose two major tasks: first, the application of a modified response surface type model to complex psychological phenomena (this application will involve ordinal-scaled or integer-valued variables) and second, within the context of this response surface model, the investigation of a number of important independent variables as they relate to a classroom concept learning task. The procedures will attempt to avoid the criticisms which have been directed toward most previous work in this area, specifically, the artificiality of the experimental task, the finiteness of the system, the lack of a vocabulary problem, and the immediateness of the measure of learning.

The specific variables to be studied include:

1) amount of redundant information
2) mode of presentation of successive instances
3) ratio of positive and negative instances
4) order of positive and negative instances
5) amount of information in accompanying verbal cues
6) length of time the instances are available to subject
7) length of time between instances
8) relative complexity of concept.

Outline of the Remaining Chapters

Chapter Two contains a brief survey of the response surface methodology literature, an exposition of the mathematical model which is the basis of response surface methodology, and a discussion of the difficulties in applying the model to experimental problems in education.

Chapter Three is a description of the concept learning experiment. Some findings from the literature are reviewed, the rationale for the study presented and the experimental procedures described.

The data collected from the set of experiments will be summarized in Chapter Four. The results will be discussed first from the experimental design viewpoint and, secondly, as they contribute to the concept learning area.

Interpretation of the results will appear in Chapter Five along with a statement of conclusions reached and recommendations for future use of the findings of this study.

CHAPTER II

REVIEW OF RESPONSE SURFACE METHODOLOGY

This chapter will briefly survey the Response Surface Methodology (RSM) literature, present in detail the mathematical model, and discuss the various strategies to be used in applying the techniques to the exploration and estimation of a response surface. The problems encountered in using RSM with discrete and ordinal scaled variables will be discussed and an approximate empirical solution to these difficulties as they apply to the exploration phase will be suggested.

Survey of Response Surface Methodology

RSM is a comprehensive experimental strategy for optimizing responses, seen as functions of sets of independent variables, and for estimating the functional relationship between response and variables that exists in the region of the optimal response. In its essence RSM is an integration of least squares--regression theory, geometric interpretation of algebraic equations, and empirical optimization strategies for use with sequential experimentation. While parts of these topics have been

well understood for a long time,* it was not until 1951 that G.E.P. Box and K. B. Wilson (1951) put them together and coined the term, response surface. In the intervening years Box and others (Box and Hunter, 1957; Box and Draper, 1959) have modified the strategy and techniques, primarily through the establishment of criteria for choosing experimental designs and the development of second- or higher order designs. In formulating practical approaches to the application of RSM within an ongoing industrial production system, Box (1957) has coined another term, evolutionary operation, to indicate operational procedures leading to experimentally determined optimal responses within ongoing full-scale production runs.

An early examination of the possibilities of experimentally optimizing a response assumed to be some unknown function of independent variables and subject to error was by Hotelling (1941), who directed himself to the problem of finding the maximum response to an unknown function of a single variable. In so doing he suggested procedures and raised questions which anticipated the work of Box during the 1950's. Hotelling

---

*Newman (1956) points out that Gauss and Legendre independently formulated the theory of least squares during the first decade of the nineteenth century and he credits Descartes and Fermat with "fathering" analytic geometry during the 1600's.

suggested fitting a quadratic polynomial to observations taken as chosen values of the independent factor or variable and noted that two sources of error must be considered: that due to inaccuracy in taking observations and that due to bias resulting from the fact that the unknown function might not be quadratic. Plackett and Burman (1946) addressed themselves to the problem of the effect that alterations in components of a manufacturing process would have upon some measured characteristic of the final assembly. While concentrating on static designs, primarily for the situation in which each component has only two possible values, their applications of least squares theory and criteria for optimal multifactor designs definitely presage the later work of Box.

The theoretical development of RSM has been largely the work of Box and various associates and is contained primarily in three basic papers. Box and Wilson (1951) present a thorough statement of the model, the underlying philosophy, a criterion for choosing RSM designs, and examples of the use of RSM in the chemical industry. In the second paper, Box and Hunter (1957) introduce the concept of the variance function for an experimental design and advance reasons for preferring designs which are rotatable. First- and second-order designs meeting the criterion of rotability are given. The third paper,

by Box and Draper (1959) introduces the bias criterion for design selection. The content of these three papers will be discussed more fully in the section on the RSM model.

In addition to these three papers several good, general statements of RSM have appeared. Box (1954) presents a thorough discussion of the principles underlying the method with only limited explanation of calculation procedures. A full explanation of the model, methods, and calculation procedures, as well as a statement of the basic principles and strategy is in Chapter 11 of The Design and Analysis of Industrial Experiments edited by Davies (1956) and co-authored by Box. Box and Hunter (1958) present another thorough presentation in a chapter in Experimental Designs in Industry edited by Chew. A somewhat less detailed, but very readable, review of RSM appears in a series of articles by Bradley (1958) and Hunter (1958, 1959a, 1959b) directed toward the industrial quality control worker.

The remainder of the theoretical literature is concentrated upon the construction of specific designs and the examinations of the statistical properties of those designs. With the exception of an article by Spendley, Hext, and Himsworth (1962) suggesting the usefulness of simplex designs to evolutionary operation procedures,

these theoretical papers tend to have little applicability to the problems considered in this paper, largely because of the restrictive mathematical properties required by the models. A very complete literature survey of RSM written by Hill and Hunter (1966) lists these peripheral theoretical papers, as well as an extensive enumeration of the variety of practical applications which have been found for RSM. As indicated in Chapter One, these applications are to a wide variety of industrial and agricultural uses, including the manufacture of truck tires (Weissert and Cundiff, 1963), growth of coastal bermuda grass (Welch, Adams and Carmon, 1963), maximization of potato yields (Hermanson, 1965), production of machine tools (Wu, 1964), and improvement of pie crusts (Smith and Rose, 1963). The only reference to RSM to the behavioral sciences that this author has found is the dissertation work of Meyer (1961, 1963) which is theoretical rather than applied.

Meyer (1961) investigated the use of RSM with integer-valued factors but the emphasis was upon the estimation of the shape of that surface in some static region rather than upon the exploration in search of an optimal region. By Monte Carlo procedures he examined a long series of designs with integer-valued factors and compared them to corresponding optimal designs with continuous-valued factors.

In summary, the theoretical model for use with continuous variables is well developed and applications to a wide variety of industrial situations have been reported. However, applications of the RSM model to empirical problems in education and psychology seem to be nonexistent. While this condition is due in large part to the restrictions which the mathematical model places upon the nature of the independent variables and the metric in which they are measured, it is probably also true that lack of familiarity with the technique on the part of behavioral researchers has contributed to this void of behavioral applications.

The Response Surface Model

The general problem for which RSM was developed is that of empirically investigating a workable production system for the purpose of identifying, if possible, combinations of the input variables to the system which would optimize some measure of the yield or output of the system. In the discussion that follows the inputs will be referred to as variables and the outputs as responses. The responses may be measures of quantity or quality of output, in which case the interest is in identifying combinations producing a maximum response, or it may be a measure, such as cost per unit produced, in which case the optimal response would be a minimal one.

That the system is a workable one implies that a combination of relevant variables has been identified which results in a minimally acceptable response. Whether the response is one, such as amount of chemical produced seen as a function of time, temperature and per cent solution of catylyst, or grain yield per acre as a function of amount of fertilizer and irrigation, or amount of learning as a function of number of examples and time spent on each, it is highly unlikely that the original combination which results in an acceptable product will be close to the optimal combination. Thus, the first aspect of RSM is the development of experimental procedures for efficiently moving from some original workable combination of variables to a combination which will, in some sense, result in an optimal response.

Once one has identified what appears to be an optimal response on one criterion, it is reasonable to investigate the possibility of multiple optimal combinations and, should they exist, attempt to identify among them that combination which is optimal on some second criterion. For example, if a number of combinations of levels of variables result in equal maximal quality of product, one might inquire as to which of these minimizes the cost per unit produced. This leads to the second aspect of RSM, that of identifying experimental

designs which will enable one to efficiently estimate the nature of the relationship that exists between the variables and the response when the levels of the variables are in the optimal range. Thus, RSM has two aspects, the first exploratory and the second aimed at the estimation of the functional relationship.

Box and Wilson* (1951) originally stated the problem as follows: One is interested in a response, η, which is supposed dependent upon the levels of a set of k quantitative factors or variables, $X_1, \cdots, X_i, \cdots, X_k$. These variables are subject to exact measurement and control. Thus, for the u^{th} combination of the levels of the variables $(u=1,2,\cdots,n)$,

$$\eta_u = f(X_{1u}, X_{2u}, \cdots, X_{ku}).$$

This unknown function, **f**, will be referred to as the response function. The observed response to any combination of levels of the variables, y_u, is subject to experimental error and, in repeated observations, varies with mean η and variance σ^2.

Viewed geometrically, the response and the k variables define a (k+1)-dimensional Euclidean space. The response to any given combination of levels of the

---

*This description of the model, experimental strategy, and criteria for choosing RSM designs is based upon the work of G.E.P. Box and his associates, particularly the basic articles mentioned on page 17, Box and Wilson (1951), Box and Hunter (1957), and Box and Draper (1959).

variables is a point in that space and the totality of points corresponding to the responses to all possible combinations of the variables form a surface over the k-dimensional hyperspace defined by the variables. This surface is the response surface (RS).

Circumstances which will influence the strategy of the investigator in exploring this RS are the magnitude of the experimental error, the complexity of the response function (or surface), and whether or not the experiments may be run sequentially with each experiment designed using knowledge from preceding ones.

One approach to identifying optimal conditions would be to examine the entire domain of the k variables. This can always be accomplished by saturating the domain with a sufficient grid of points to allow adequate approximation of the response function, f. This, in general, would require an impracticably large number of points. However, in those situations where the experimental error is small and the experiments can be run sequentially, there are strategies which will locate optimal combinations with greatly reduced experimental effort.

Small experimental error implies that small changes in η can be accurately detected and small sub-regions of the RS can be investigated with a small number of

experimental points. When experiments can be run sequentially, knowledge from one set of experiments can be used to identify a new set of conditions resulting in a higher (or lower) level of response. By a series of such moves one should move from any arbitrary beginning set of experimental conditions to a set which results in uniformly high (or low) responses, i.e., to a region of optimal response. While this region may only provide a local rather than an ultimate optimum, this is an inevitable risk of limited exploration of the variable space.

While the entire surface (or its algebraic equivalent, the function f) may be quite complex, knowledge of many phenomena suggest that a reasonable assumption is that it is both smooth and continuous. Thus, the function may be approximated* to any desired degree of accuracy by a polynomial of sufficiently high degree. If either k, the number of variables, or d, the required degree of the polynomial, is large a polynomial of the form

$$\begin{aligned} p = {} & a_0+a_1X_1+a_2X_2+ \cdots +a_kX_k \\ & +a_{11}X_1{}^2+a_{22}X_2{}^2+ \cdots +a_{kk}X_k{}^2+a_{12}X_1X_2+ \cdots \\ & +a_{111}X_1{}^3+a_{222}X_2{}^3+ \cdots +a_{kkk}X_k{}^3+a_{112}X_1{}^2X_2+ \cdots \end{aligned}$$

---

*The Weierstrass approximation theorem states: If f is continuous on the closed interval [a,b], then it can be uniformly approximated to within ϵ on [a,b] by a polynomial, for any $\epsilon>0$. (Buck, 1956; p. 39)

can require the estimation of a considerable number of constants.* The number of constants to be estimated for $k=1,2,\cdots,5$ and $d=1,2,\cdots,5$ are given in Table 1. Since the minimum number of points needed to fit any polynomial is the number of constants, the figures in Table 1 present the minimal numbers of experiments needed to fit the approximating polynomial for given d and k. Add to

TABLE 1

Numbers of constants to be estimated in fitting a polynomial of degree, d, in k variables.

Number of Variables (k)	Degree of polynomial (d) 1	2	3	4	5
1	2	3	4	5	6
2	3	6	10	15	21
3	4	10	20	35	56
4	5	15	35	70	126
5	6	21	56	126	252

this minimal number the experiments needed to adequately sample the entire domain of the variables and the replications necessary to estimate σ^2 and the experimental effort needed to adequately approximate f over the entire domain of the X_i is, in general, impossible in practice.

Fortunately, in most applied situations practical restrictions on the variables limit the domain which is

*The notation used in writing polynomials subscripts each constant with a set of indices consisting of all the indices of the corresponding variable or product of variables. Thus, the coefficient of $X_2^2X_3=X_2\cdot X_2\cdot X_3$ is a_{223}.

of initial interest and the region of ultimate interest, that over which one would like to estimate **f**, is even further restricted to areas of optimal or, at least, near-stationary response. The result of these restrictions on the domain of interest is that polynomials of degree 1 or 2 usually provide adequate local approximation to **f** bringing the needed expenditure of experimental effort within reasonable practical levels.

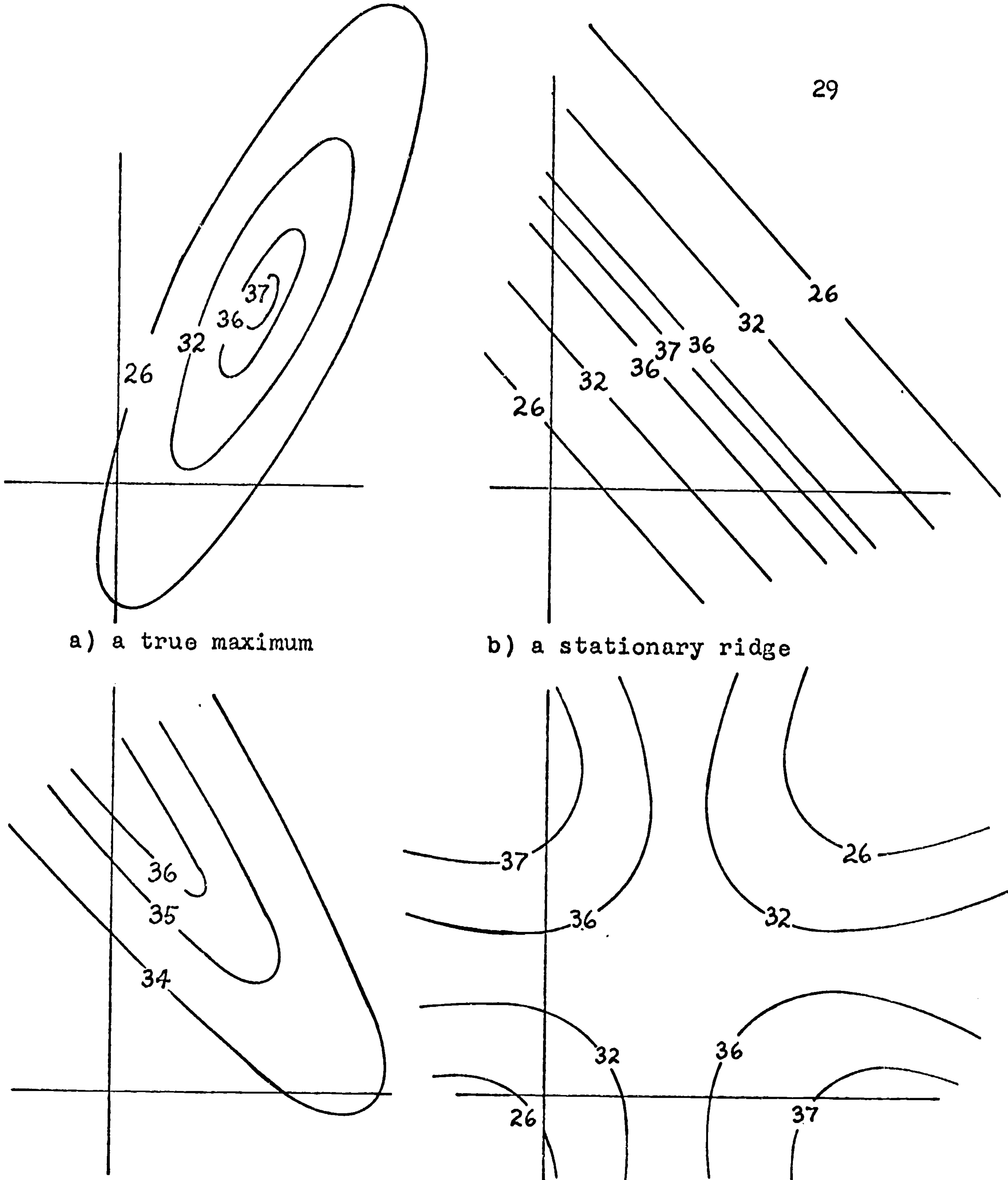
One begins the investigation of a RS in some arbitrary sub-region of the variable space. It is unlikely that responses in that region are either minimal or maximal. If the region is not one in which the response is near optimal, then one is in a region where the RS has a slope relative to the hyperplane defined by the variables. The farther one is from either a maximum or minimum, the greater that slope is likely to be.

The first phase of the RS problem is that of efficiently moving from this arbitrary initial region of investigation to an optimal one. Box suggests that the move should be that which results in the greatest increment in the response, i.e., perpendicular to the contour lines of the RS or in the direction of steepest ascent. Since the only information desired from the initial set of experiments is that of which direction to move next, and, since the slope of the RS is likely to be relatively steep, the local approximation of the

RS by a plane should be sufficient. As one, through a sequence of moves, gets closer to the optimal region the slope will decrease and a quadratic or cubic surface may be needed.

The local approximation of the RS by a plane is the equivalent of estimating f by the first degree polynomial, $p_1=a_0+a_1X_1+\cdots+a_iX_i+\cdots+a_kX_k$. The direction of steepest ascent can be found by calculating the first order partial derivatives, $\frac{\partial p_1}{\partial X_i}$, of p_1. The direction of steepest ascent will be that taken when increments of the X_i are proportional to these partial derivatives. A check on the adequacy of the calculations is provided by the extent to which the indicated increments of X_i do lead to increments in the response.

Movement is made to a sub-region up the slope of the RS from the initial region, a new set of experiments run, and the slopes redetermined. By a series of such steps one moves to regions of higher and higher response. However, the higher up the RS one moves, the more gradual the slope of a plane becomes and the more likely one is to move as a result of experimental error rather than a true increment in response. Also, the plane becomes a less and less satisfactory approximation to the RS. One is in a near-stationary region. The second phase of RSM is the exploration of this region.



a) a true maximum

b) a stationary ridge

c) a slowly rising ridge

d) a minimax or saddle point

FIGURE 1

Typical Response Surfaces in regions
of near-stationary response.

This second phase of the investigation attempts to discover whether the near-stationary region is a) a true maximum (as in Figure 1,a), b) a ridge system, either stationary or slowly rising (as in Figure 1,b and c) or c) a saddle point or col or minimax (as in Figure 1,d). If the region is not too large, a second-degree polynomial should provide an adequate approximation.

The designs suggested by Box for this first phase exploration are the two level factorial designs and, in situations where it may be assumed that factors will not interact, the two level fractional factorials. The center of each design is chosen as the origin and the units of each variable so chosen that the levels of each are +1 and -1.

The choice of appropriate units for the several variables needs comment as the slopes of the fitted plane and resulting path of steepest ascent are clearly not invariant under a change of measurement scale. The procedure which makes most sense for choosing appropriate units for the variables is that of choosing units so that unit increments in any single variable lead to approximately similar increments in the response. Success in these choices are obviously a function of the skill and experience of the experimenter. Adept choice of scale

results in an approximately symmetrical representation of the response function which reduces the amount of calculations to be done in the estimation phase of the problem. Because of the sequential nature of the procedures, an inappropriate choice of scale unit may be detected and corrected in later experiments.

When any one variable produces effects that are quite small compared to those of other variables, it may be that:

a) the unit chosen is inappropriately small,

b) the response is independent of that variable, or

c) one is at a conditional maximum for that variable.

By increasing the size of the unit and changing the level of that variable from the calculated path, one may determine the exact situation.

The approximation of the RS in the near-stationary region requires a polynomial of, at least, degree two. An experimental design must be chosen which will both provide efficient estimates of the constants in the polynomial and facilitate the recognition of the characteristics of the RS in that region. Given an appropriate design, a second-degree polynomial can be fitted to the RS. A check on the adequacy of the fit can be made by partitioning the sum of squares into that due to the approximating polynomial (due to regression) and that

which represents lack of fit (due to deviations from regression). An analysis of variance may be done and the mean square for lack of fit examined. If the second-degree polynomial provides an adequate approximation to the RS, this mean square for lack of fit is an estimate of σ^2, the variance due to experimental error. If the fit is not adequate then the mean square should be inflated by amounts due to the extra constants and should be considerably larger than σ^2. Thus, if an independent estimate of σ^2 is available, a comparison of this estimate with the mean square provides a test of the adequacy of the second degree polynomial in approximating the RS.

Given that an adequate approximating polynomial has been calculated, one still has the problem of recognizing the true nature of the surface from the study of the equation. As the polynomial will be oriented about an arbitrary origin, it will, in general, be difficult to learn much about the surface from direct examination to the constants. Box suggests that the polynomial be transformed to canonical form, i.e., to a form oriented about an origin at the estimated point of optimal response with axes along or perpendicular to the major axis of the RS.

The canonical form for a second degree polynomial is of the form

$$Y - Y_m = A_{11}x_1^2 + A_{22}x_2^2 + \cdots + A_{kk}x_k^2$$

where Y is the response, Y_m the response at the estimated point of maximal response, and the x_i the set of axes corresponding to the axes of the RS.

The estimated point of optimal response is found by taking the partial derivitives of the approximating polynomial relative to the k variables, X_i, setting each equal to zero and solving for the coordinates of the point. The A_{ii} may be found by solving the characteristic equation

$$\begin{vmatrix} a_{11} & \tfrac{1}{2}a_{12} & \cdots & \tfrac{1}{2}a_{1k} \\ \tfrac{1}{2}a_{12} & a_{22} & \cdots & \tfrac{1}{2}a_{2k} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \tfrac{1}{2}a_{1k} & a_{2k} & \cdots & a_{kk} \end{vmatrix} - \begin{vmatrix} A_{11} & 0 & \cdots & 0 \\ 0 & A_{22} & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & \cdots & A_{kk} \end{vmatrix} = 0$$

for the A_{ii}. The transformation from the original X_i system to the x_1 one oriented to the RS is accomplished by solving the matrix equation

$$\underset{k\times k}{\begin{bmatrix} T \end{bmatrix}} \underset{k\times 1}{\begin{bmatrix} X_i - X_{im} \end{bmatrix}} = \underset{k\times 1}{\begin{bmatrix} 0 \end{bmatrix}} \quad \text{for } \begin{bmatrix} T \end{bmatrix}$$

subject to the restriction that $\begin{bmatrix} T \end{bmatrix} \begin{bmatrix} T' \end{bmatrix} = \begin{bmatrix} I \end{bmatrix}$ where $\begin{bmatrix} I \end{bmatrix}$ is the k k identity matrix.

In the case where there were only two variables the second degree canonical form would be

$$Y - Y_m = A_{11}x_1^2 + A_{22}x_2^2.$$

Examining the possible shapes of the surfaces pictured in Figure 1, the advantages of the canonical form can

be easily seen. When the contour lines of the surface are ellipses (as in Figure 1, a) then both A_{11} and A_{22} will be negative with the relative absolute magnitude of the constants indicating the degree of elongation. The presence of a ridge system (as in Figure 1, b) would be indicated by one of constants being approximately zero. Should the ridge be of the slowly rising type, indicating a maximum at infinity (as in Figure 1,c), the origin could be taken at some convenient point on the axis of the ridge, in which case the canonical form would be

$$Y - Y_m = A_1x_1 + A_{11}x_1^2$$

with A_1 being the slope along the x_1 axis. When the signs of the constants, A_{11} and A_{22}, are opposite one has a RS in the shape of a saddle point or minimax (as in Figure 1,d).

To this point the need to approximate the RS by a polynomial has been indicated several times, but nothing has been said of the procedure for calculating the best polynomial. The actual procedure used is that of fitting the polynomial to the RS by the procedures of linear least squares or curvi-linear regression. That is, the constants, a_i, in the polynomial are chosen so as to minimize the sum of squares of the deviations between the responses predicted by the equation and those actually observed as a result of experiments.

The basic assumption of least squares is that the observed response is a linear function of a set of latent parameters or constants. It should be noted that this permits the estimation of polynomials since the linear restriction is on the constants, not the variables. Further, while the variables are assumed error-free this does not seem to be a particularly serious source of difficulty when the variables are levels or powers of levels of factors which are set prior to the running of experiments as they are here.

In the previous pages the overall strategy and general procedures for a response surface exploration and estimation problem have been outlined. The next problem to be considered, and that which has monopolized the efforts of theoreticians in this area is the experimental design problem of carrying out these procedures in the most efficient manner. That is, what strategies can be devised for selecting experimental combinations of levels of the variables that will provide the greatest amount of information, i.e., the best fitting polynomial, for a given amount of experimental effort. The three major contributions to this problem have been the original Box and Wilson (1951) paper which introduced central composite designs, Box and Hunter's (1957) suggestion that desirable designs have the property of

rotatability and the Box and Draper (1959) paper emphasizing the importance of the criterion of minimal bias.

Box and Wilson (1951) point out that, while the two level factorial and fractional factorial designs are adequate for estimating plane surfaces, the efficient estimation of quadratic and cubic surfaces requires a different configuration of points. The composite designs suggested are basically factorial designs which may be augmented by additional points at the center and on the axes of the design should a first degree polynomial be deemed an inadequate fit.

Box and Hunter (1957) introduced the concept of the "variance function," V(x), of an experimental design, which they define as

$$V(x) = N\ V(\hat{y}_x)/\sigma^2$$

where $V(\hat{y}_x)$ is the variance of the estimated response at any given point. The variance function provides a standardized measure of the precision with which the design estimates the surface. If the value of V(x) at any given point is strictly a function of the distance of the point from the center of the design then the design is said to be rotatable, i.e., rotation of the axes, as in the transformation to canonical form, will have no effect on the precision with which the design estimates the surface. In the situation where one does not

know the orientation of the RS to the hyperplane of the variables prior to the running of the experiment there are obvious advantages to a design meeting this criterion.

However, Box and Draper (1959) point out that in approximating any unknown function by fitting a polynomial to empirical data one not only runs the risk of error due to sampling, but, also, due to possible inadequacies in the polynomial as a representation of the true function. This latter source of error is referred to as bias error, in contrast to the former, variance error. In investigating the problem of designs to minimize both sources of error, they studied the situation where a first degree polynomial was used to approximate a function that was truly a second degree polynomial. They arrived at the conclusion that, when both variance and bias error are present, the optimal designs for minimizing both sources of error are very similar to those which ignore variance error completely and only minimize bias error. This is particularly important since the two criterion, minimize variance error and minimize bias error, lead to different design principles. Minimizing variance error leads, in general, to large designs with points outside the region of immediate interest while the bias criterion calls for designs with experimental points within the region of immediate interest.

The additional theoretical papers which have appeared since the Box and Draper paper have attempted to apply these ideas to more complex situations where the approximating polynomial and the true function are of higher degree, but no really new approaches or strategies have been proposed.

An interesting extension of RSM has been to the problem of maintaining an ongoing production process at near-optimal conditions in spite of the effects of changes in the quality of input materials, systematic changes in instrumentation, and differences between the conditions of a full scale production run and those of a limited, more tightly controlled laboratory routine. Since all production processes have tolerance limits within which the quality of product is acceptable, it is possible to introduce small variations in the production conditions leading to information about the continued optimal characteristics of those production conditions while still maintaining acceptable product output. Box (1957) introduced this idea under the name, evolutionary operation (EVOP). EVOP is, in essence, the continual application of the optimization or exploration phase of RSM to an ongoing production process. The basic idea underlying EVOP is that the operation of any industrial process for the sole purpose of producing a product is inefficient;

the process should be made to produce both product and information as to how the product might be improved. Since one is only concerned with the levels of variables which produce optimal responses the local approximation of the RS by a plane is adequate and the design for the variations in the "works process" can be quite simple. Box (1957), in explaining EVOP, uses a 2-level factorial.

Another design which has many desirable properties for investigating a RS when a plane is deemed a sufficient approximation is a simplex design. While Box and Hunter (1957) only allude to the desirable characteristics of simplex designs, a full description of the application of the design to optimization and EVOP problems is presented by Spendley, Hext, and Himsworth (1962). Since it is the use of a simplex design which is followed in the sequence of concept learning experiments to be described in Chapter Three, the usefulness of the design, as presented by Spendley _et al._, will be described at length.

A simplex is that minimal geometric shape which can exist in a Euclidean space of given dimension but not in a space of lesser dimension. Thus, a line segment is a simplex in one dimension, a triangle a simplex in two dimensions and a tetrahedron a simplex in

three dimensions with the concept generalizable to k-dimensional space. A regular simplex is one in which the distances between vertices are equidistant. In an experimental situation with k variables, a regular simplex design would consist of (k+1) experimental combinations representing the vertices of a regular simplex.

The basic simplex design with k variables is given by

$$D= \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ p & q & q & \cdots & q \\ q & p & q & \cdots & q \\ q & q & p & \cdots & q \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ q & q & q & \cdots & p \end{bmatrix}$$

where $p = \frac{1}{k\sqrt{2}}\left\{(k-1)+\sqrt{k+1}\right\}$ and $q = \frac{1}{k\sqrt{2}}\left\{\sqrt{k+1} - 1\right\}$.

Each of the (k+1) rows of this matrix are the coordinates of a vertex of the simplex. Regularity is not seen as a problem when the variables are measureable on continuous, interval scales, because regularity can easily be obtained by an appropriate linear transformation of any one or more of the variables.

At this point certain characteristics of a simplex should be noted. A simplex of any given dimension, say k, has (k+1) vertices and (k+1) faces. Each face of a k-dimensional simplex is a simplex of (k-1) dimensions. The removal of any one point or vertex from the design, thus, collapses the design into a simplex of one less dimension. Correspondingly, the addition of a new point

to the design (a point not on the hyperplane defined by the existing simplex) creates a new simplex of one larger dimension.

Since the exploration of a RS is a series of moves of an experimental design from one region of the variable space to another, the characteristic of a simplex design which permits one to create a new simplex by removal of only one vertex and the subsequent substitution of any new point not on the hyperplane defined by the remaining points is a particularly useful one.

Any exploration procedure consisting of a set of moves requires a set of rules for making those moves. Rules are needed for decisions specifically concerned with: when to move, in what direction and how far to move, and how to calculate the new combination or combinations. The rules given by Spendley *et al.* are particularly simple.

The decision rule concerning when to move, i.e., when to change the levels of the independent variables, is that one moves or changes combinations as often as possible. Thus, one runs experiments at the first (k+1) combinations and from then on each additional experiment is a new combination, the exact nature of which is determined by the results from the immediately previous (k+1) experiments.

Two logical alternatives in a rule for deciding in which direction to move would be either to move away from the point of least optimal response or to move toward the point of most optimal response. Note that both of these procedures divide the original set of (k+1) points into two subsets of more or less optimal responses, one of 1 point and one of k points. The first alternative defines the least optimal subset as having a single point while the second defines this subset as having k points. Since the purpose is to find a region of optimal response, the changes made should be those of replacing the least optimal set with a new set of points which lie on the opposite side of the most optimal set from those being rejected. Thus the implication of the first alternative is the replacement of one point by another at each step of the sequence while the second implies that a set of k least optimal points be replaced by a new set of k points at each step. In either case the new configuration of points should approximate the mirror image of the original configuration.

The second alternative clearly results in the greatest change in position of the simplex design but also requires the greatest amount of experimental effort, in the two variables case, twice as much. Furthermore, since the responses at all experimental points are

equally subject to error, the second rule clearly involves the most risk as it involves using the information from (k+1) responses to make k changes while the first approach is using the information from the (k+1) responses to make a single change. On the basis that the risk of incorrectly responding to responses subject to error should be minimized, the first alternative is followed, i.e., reject the single least optimal point and choose a new point opposite the face of the simplex defined by the remaining k points. By keeping the simplexes of equal size, one also controls the distance moved.

The third decision is that of determining the levels of the variables for the new combination. That point which will result in a new simplex on the opposite side of the face defined by the k more optimal points is found by taking, as any coordinate of the new point, twice the average of the corresponding coordinates of the remaining points minus the corresponding coordinate of the discarded point.

Several advantages of the use of simplex designs with these decision rules should be noted.

1) The arithmetic involved, both in the calculation of new design points and in the evaluation of the response measures, is relatively simple compared to the steepest ascent methods

suggested by Box and Wilson (1951).

2) The necessary decision rules can be simply stated and easily followed. Neither hypothesis testing nor interval estimation procedures are used, the idea being that a decision inappropriately made due to errors in the responses will be corrected in subsequent steps of the procedure.

3) The direction of advance of the design from the original set of points to a set in the region of optimal response is dependent solely on the rank order of the responses and not on their value on any absolute scale. Thus the response data may be judgments or rankings rather than measurements in the interval or ratio scale sense of the term.

4) One need depend little on any assumption of planarity of the response surface except in the immediate region of any single simplex, an assumption which is quite tenable given appropriate choice of the levels of the various factors.

5) The dimensionality of the variable space can readily be expanded or contracted. A new factor or variable, previously held constant, can be added to the design by the addition of

a single new point. The contraction of a design, by removal of a variable which appears to be having little effect on the response, is somewhat more complicated because it could result in loss of regularity.

As in any system where experimental error is present, one will occasionally make a false move, i.e., a move in an inappropriate direction, because the treatment effects are obscured by the error. This is not particularly serious, however, as the procedure calls for continuous review of past decisions. Errors at one step of the sequence can be expected to be caught at a later stage and the greater the adverse effect of the inappropriate move, the more rapidly it will be detected.

The size of the error standard deviation can be reduced by replication of the experiments at each combination but Spendley et al. demonstrates that it is more efficient not to expend effort in replication but to instead run a longer sequence of experiments. He also presents empirical data that suggests that the procedure works better with larger numbers of variables than with smaller.

Problems of Applying RSM to Educational Research

The major premise upon which scientific, meaning empirical and inductive, investigation is based is that

of the essential lawfulness of nature. The outputs occurring in any natural system or subsystem of interest must be viewed as functionally dependent upon both the initial state of the system and the kinds and amounts of relevant inputs to the system. Thus, in the behavioral sciences, as well as in the natural sciences, the ultimate outcome of research activity should be the specification of the relationships which exist between status and input variables on the one hand and outputs on the other. However, when the variables are categorical rather than quantitative, ordinal rather than interval, or discrete rather than continuous, as they often are in the behavioral sciences, difficulties with the mathematics needed to construct an appropriate model have often forced researchers to adopt experimental or research strategies which obscure this ultimate goal of the specification of functional relationships. Even in areas where the existence of relationships among sets of variables are well established and the nature of those relationships indicated, researchers in education still revert to procedures which can only answer the preliminary question concerning what variables are related or the comparative question inquiring as to which of two, possibly mediocre, choices is the better. In contrast, RSM begins with the assumption that the gross nature of the relationships among the variables in a

system is known and aims at answers to the questions of which is the best combination of variables and what is the precise nature of the relationships that exist.

It is the contention of this writer that there are areas of education where the gross nature of relationships among variables are well enough understood to warrant an attempt to fit this knowledge into the RSM model that has been described and attempt both the optimization of responses and the approximation of relationships existing between variables and response in the region of optimal response. While the second task is, perhaps, the more difficult, the first must take precedence and the empirical aspects of this paper will attempt to demonstrate that this can be done.

The difficulties in applying RSM to the problems of education are several. Probably the most apparent difficulty is with the metric characteristics of the variables identified as relevant. All too often these are categorical rather than quantitative. Since categorical implies that there is no rationale for ordering the levels of the variable, there can be no RS in the sense discussed here. However, researchers in education have too willingly accepted identified categorical variables as important for study and have, perhaps, spent too little effort in attempting to analyze categorical variables into possible underlying quantitative variables.

Another metric problem is that of measuring those variables which are quantitative on scales which have the relevant metric properties--those of being continuous and equal interval. Neither is an insurmountable obstacle if one is willing to tolerate a certain amount of imprecision and inefficiency in the design. The use of a scale which is ordinal but not sufficiently equal interval will cause the response function to be unnecessarily complex but should not seriously interfere with the optimization process. When the variable is discrete, the RS breaks down and becomes a lattice or net of points rather than a surface. While it may make good experimental sense to operate as though this lattice or net of points would all lie on some smooth surface, one should realize that the mathematics upon which RSM is built requires that the totality of responses form a surface and that the violation of this requirement could lead to peculiar results.

The changing nature of the experimental material is yet another difficulty. Most of the responses of interest to educators are some concomitant of physical or cognitive growth and any procedure which runs experiments consecutively rather than simultaneously runs the risk of confounding treatment effects with those of normal growth or development.

γ

Because RSM dictates that experiments be run sequentially, one must be especially careful that the running of an earlier experiment doesn't contaminate the experimental material to be used later. Combined with the possibility that a long sequence of experiments may be run, one needs access to a relatively large supply of subjects.

It should be noted that the difficulties with the mathematics due to the grossness of measurement techniques is far more serious for the problem of estimation of relationship in the area of optimal response than it is for the exploration phase leading to that optimal response.

Summary

In surveying the RSM literature, the pre-eminent role of G.E.P. Box in its theoretical development was stressed. An indication of the breadth of practical applications of the methodology in industry was contrasted to the void of application in behavioral research.

The mathematical model has been presented in some detail, both the exploration and estimation phases. An approach to the exploration phase due to Spendley _et al._ and utilizing certain very desirable properties of simplex designs has been outlined.

Difficulties in applying RSM to behavioral, specifically educational, research have been discussed and

the point made that the difficulties are far more serious with the estimation phase of the procedure than with the exploration.

CHAPTER III

AN APPLICATION TO CONCEPT LEARNING

As indicated in Chapter I a major purpose of this dissertation was to demonstrate that the experimental procedures and strategies presented in Chapter II could be usefully applied to areas of behavioral science of specific interest to education. This chapter is a description of such an application. A brief introduction to the topic of concept learning and the rationale for using it as the vehicle for testing the optimization ideas of RSM will be presented first. This will be followed by a review of the procedures followed in running the sequence of experiments including those dealing with development of the experimental task, the response measure, the population from which subjects were drawn and the steps followed in the actual performance of the experiment.

As indicated at the conclusion of Chapter II, the basic requirement for the conceptualization of a research problem in the RSM pattern is that the variables and response of interest be quantitative. A second requirement is that a supply of relatively stable experimental material be available. Third, one must be working within an experimental procedure which will permit

rather strict control of irrelevent variables and will yield a relatively small standard error. Finally, a research area in which a number of relevent variables had been identified would permit an examination of the methodology in a truly complex setting. The topic of concept learning seemed to fit this bill of particulars.

Review of Concept Learning

Interest in the study of the psychological process of conceptualization has been of concern to psychologists since the beginning of the century under such names as generalizing abstraction (Fisher, 1916), concept formation (Hull, 1920; Smoke, 1932), concept attainment (Heidbreder, 1946), concept learning (Hovland, 1952), and concept identification (Archer *et al.*, 1955) Examination of the experimental tasks required of subjects by all of these investigations reveals sufficient commonality to conclude that all were interested in aspects of a single psychological process which will be referred to as concept learning.

The chronological trail from the early work of Moore (1910) and Fisher (1916), who relied exclusively on introspective reports of subjects for data, to the first attempts at quantifying concept learning by Hull (1920), to the relatively recent work of Bruner, Goodnow, and Austin (1956) reveals two interesting trends. The

most obvious one is that of moving towards greater objectivity by repeated modification of the experimental task in the direction of removing equivocal sources of variability. The second trend, which is actually a result of the first, is that of the experimental task becoming a less and less adequate analog of what a classroom teacher would consider the learning of a concept. The subject is being asked to solve a finite problem rather than add a new concept to his cognitive repertoire.

The early experiments of Fisher, Hull, and Smoke required the learning of arbitrary concepts which were characterized by having a novel, nonsense name and by the possibility of their being represented by an infinite number of percepually differentiable exemplars. By contrast, examination of the recent work of Bruner *et al.*, which has set the pattern for much of the experimental work of the past dozen years, leads to the characterization of the concepts learned as having a name familiar to the subjects, being applicable to a finite set of stimuli and having little or no permanence. In the introductory remarks in Chapter One a case was made that tasks presently being used in concept learning studies were inadequate analogs of tasks presented children by teachers under the rubric of learning concepts. The

points of inadequacy were the finiteness of the universe of exemplars of the concept, the use of familiar vocabulary in familiar ways to name the concept, the absence of any expectation that the learning will persist and the often systematic way in which stimuli are presented.

The chronicle of this transition from tasks which were quite adequate analogs of real life concept learning to present tasks which are more closely analogous to the solving of a problem is long, devious and obscure. The obscurity of the trend is, in no small measure, attributable to a tendency on the part of researchers to avoid defining ambiguous terms. Perhaps a single example from the very influential _A Study of Thinking_ (Bruner _et al._, 1956) will demonstrate what seems to have happened in the literature on this topic.

Bruner _et al._ develop their idea of how concept attainment takes place by describing the manner in which a hypothetical foreigner might, over a period of time, by using information about individuals that he meets, develop and attain a concept of what an "influential" person is. The process described is one of meeting individuals, noting characteristics of the individual and receiving information as to how influential the person is. Bruner _et al._ maintain that by a variety of strategies this information is processed and results in a concept,

in this case that of an influential person. This discussion of the theory underlying the attainment of a concept is concluded by an unsubstantiated statement, "One can specify the minimum array of instances necessary in order for our hypothetical foreigner to solve the problem of who is influential." (Bruner et al., 1956; p. 65). No evidence is presented that such an array can be specified and there is no recognition by the authors that a very basic assumption has been made. This statement implies that the list of relevant attributes of the concept, influential, is both finite and determinate. If one can accept this assumption, then the use of finite arrays of geometric figures as the stimulus material for concept learning tasks can be accepted. However, this writer cannot accept this assumption and, therefore, cannot accept "blue circles" as an equivalent type of cognitive thing to "influential person." Yet this obscure statement does serve as a connecting link between the theoretical and the empirical portions of the Bruner book.

The experimental papers which have appeared in the educational and psychological journals during the past fifteen years have, by and large, appeared to make the same assumption. Article after article (e.g., Hunt and Hovland, 1960; Glanzer, Huttenlocher, and Clark, 1963;

Kates and Yudin, 1964) purports to be studying some aspect of concept learning with the only justification for the product of the experimental task being called a concept being that others have used similar tasks to study concepts. The point of this discussion is not that the empirical evidence which has been accumulating has no relevance to the problem of how one learns a concept, but, rather, that we may have considerable difficulty transferring these findings to the classroom because certain important aspects of the concept learning task have been omitted.

The experimental procedures and tasks to be described later in this chapter have been adopted, as far as possible, to parallel the type of task used in recent empirical studies of the problem while, at the same time, meeting the objections to those tasks which have been raised here and in Chapter One.

Klausmeier, Davis, Ramsay, Fredrick, and Davies (1965) have compiled a rather complete bibliography of the concept learning and problem solving literature. Included is a taxonomy of vaiables seen as relevant to the concept learning process. Of this list of over one hundred variables, the following eight were selected for inclusion in the sequence of experiments:

1) amount of redundant information,

2) mode of presentation of successive instances,

3) ratio of positive and negative instances,

4) order of positive and negative instances,

5) amount of information in accompanying verbal cues,

6) length of time instance available to subject,

7) length of time between instances, and

8) relative complexity of concept.

The Experimental Design

The usual procedure in the design of an experiment is for the researcher to identify the problem area of interest, to formulate specific questions to be answered or hypotheses to be tested, to define the treatments to be administered, to select the population to be studied and then to build these factors into an appropriate experimental design. The problem determines the methodology.

In this situation, however, the attempt to demonstrate the practicality of the method being used was as important as the knowledge which might be gained about the manner in which children learn concepts. The order of priority was reversed and the requirements of the design dictated, within limits, the variables to be studied, the nature of the experimental task, the method of presentation, and the population to be studied.

Among the major difficulties to be faced in applying the RS model to educational or psychological problems is the paucity of important input variables that meet the criteria of being continuous, having at least an interval scale metric and being capable of exact measurement and control. What is needed are approximations to the previously described approaches which will accommodate discrete, ordinal scale variables. One such approximation, a modification of the approach suggested by Spendley, Hext, and Himsworth (1962) for the sequential use of simplex designs in optimization processes was followed in this study.

The essential difference between the design used by Spendley *et al*. and the modification used here is that they used fractional valued combinations of the various independent factors resulting in sets of points which formed the vertices of a regular simplex with edges of unit length, while in the modification proposed here one is forced to use irregular simplexes with integer valued vertices. For example, in a two factor situation, Spendley would begin with the regular simplex with vertices at the points (0.00, 0.00), (0.97, 0.26), and (0.26, 0.97), as contrasted to the irregular simplex with vertices at (0,0), (1,0) and (0,1) necessitated by discrete data.

The basic design of the modified scheme consists of (k+1) experimental combinations, where k is the number of independent variables under investigation. Relative to the chosen origin a simplex may be specified by the design matrix,

Experimental Points

$$D = \begin{array}{c} \begin{array}{ccccc} p_1 & p_2 & p_3 & & p_{k+1} \end{array} & \\ \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} & \begin{array}{c} v_1 \\ v_2 \\ v_3 \\ \\ v_k \end{array} \end{array} \quad \text{Variables}$$

where the k rows are the k independent variables and the (k+1) columns are the experimental combinations at the (k+1) points. These define a simplex in k dimensions, a simplex which forms one of the (k+1) faces of the k-dimensional simplex design.

The general experimental plan calls for the running of experiments at each of the experimental combinations represented by the columns of D, collecting response measures at each of the combinations, and then making decisions leading to more optimal experimental combinations based upon the data which have been collected from the first (k+1) experiments. The rules needed for these decisions, specifically concerned with when to move, in what direction to move and how to calculate the new combinations follow those described in Chapter Two (pages 41-3).

1. One moves, i.e., changes, the experimental combinations for each new trial after the basic (k+1) combinations have been run.
2. One moves away from the combination resulting in the least optimal response by discarding that combination from the simplex and replacing it by a new point on the opposite side of the remaining face.
3. Following the suggestion of Spendley, the procedure for calculating new combinations was as follows. For each variable, choose that integer valued level which is closest in numerical value to twice the mean level of the variable in the k combinations being retained minus the level of the variable for the point being discarded.

The aim of this calculation is to produce a new simplex in the variable space which will be a mirror image of the first. With discrete valued variables, one is forced to approximate fractional values with the nearest integer valued point. This results in each set of (k+1) points defining a simplex of somewhat different shape. This difficulty is not immediately apparent in a two dimensional problem because all mirror images of integer valued triangles will also have integer valued coordinates.

In three dimensions this is not necessarily true. The mirror image of the simplex, D_0, formed by rejecting the point (0,0,0) is D_1, with the new point, (2/3, 2/3, 2/3).

$$D_0 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad D_1 = \begin{bmatrix} 1 & 0 & 0 & 2/3 \\ 0 & 1 & 0 & 2/3 \\ 0 & 0 & 1 & 2/3 \end{bmatrix} \quad D_1{}' = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

The closest integer valued point to (2/3, 2/3, 2/3) is the point (1, 1, 1) which results in a simplex, $D_1{}'$, of slightly different size and shape than that defined by the original set of points. Furthermore, the amount of distortion that occurs from this substitution of an integer valued point for a fractional valued one is not constant and becomes potentially greater as the dimensionality of the system increases. "Potentially" must be stressed as the distortion which actually occurs will be a function of the particular sequence of simplex designs which the data collected dictates.

Description of the Variables Studied

For the purposes of this study the eight variables included in the experiment were defined as follows.

Amount of redundant information in the typical concept learning experiment refers to that information in excess of some minimal amount which the subject requires to learn the concept. The ability to specify this

minimal amount of information needed to learn the concept is, however, a consequence of the complete specificity of the set of entities used as exemplars or non-exemplars of the concept. In the real world, the number of discriminable exemplars or non-exemplars of a concept is always indeterminate and one has no way of knowing whether or not the correct categorization of some future instance will require attending to some attribute which to that time has been deemed irrelevant. This is equivalent to saying that real concepts are never complete and, thus, additional exemplars of a concept are never completely redundant in the information sense. Since it was the intention of the experimenter to build this indeterminateness into the sets of instances used, the amount of redundant information was taken to be relative and assumed proportional to the number of discriminable instances presented. Thus, a set of six instances was said to have more redundant information than a set of three even though the absolute amounts of information needed could not be specified.

Mode of presentation of successive instances refers to the context in which instances are presented. They may be presented in isolation, in conjunction with the first exemplar or positive instance, in conjunction with all previously presented exemplars or within the total

array of available instances. Note that these may be ordered in terms of the amount of prior information about the concept that is presented with each successive instance.

The ratio of positive and negative instances is a particularly interesting variable in that experimental findings point out that negative information is particularly difficult for some people to process (see Hovland and Wiese, 1953), while a logical analysis of the indeterminate real world would indicate that some negative information is necessary if one is to categorize that world in meaningful ways.

The order of positive and negative instances has received little attention in the literature. Nonetheless, it seems reasonable that it should make a difference whether instances are presented with all the positive instances first, with all the negative instances first or in some interspersed order.

This variable was included primarily because it provided added complexity for the testing of the simplex design with little added effort in the construction of treatment materials.

The use of verbal cues with varying amounts of information was suggested by the work of Wittrock et al. (1964). The cues used in this study were verbal state-

ments emphasizing certain aspects of the visual instance. While no information was presented that was not contained in the visual presentation, the higher level verbal cues operate to focus subject's attention on relevant or defining attributes of the concept.

Length of time instance was available to subject and length of time between instances were variables which were included primarily because of the ease with which they could be manipulated. Their primary function was to complicate the design and provide a more strenuous test of the simplex design and response surface model.

With finite sets of instances there is some indication that more complex concepts, i.e., those with more relevant dimensions, are more difficult to learn. As with the amount of redundant information variable, the specification of the complexity of a real world concept in any absolute sense is impossible and this variable makes sense only inasmuch as one considers the relative complexity among concepts. This is possible with the several concepts used in this study.

Experimental Task and Materials

In the introductory remarks in Chapter I, dissatisfaction was expressed concerning the artificiality of many experimental analogs of concept learning behavior and specific comments were made about three specific

deficiencies which were seen to exist in certain laboratory paradigms which are widely used in concept learning studies. These paradigms were seen to be inadequate representations of what educators would call a concept learning task in that the tasks studied are usually:

1) lacking in any requirement that the individual subject reorganize his environment in any novel way,
2) couched in vocabulary that is quite familiar to the subjects, and
3) presented in the context of a finite system of attributes having a quality of unambiguity not present in the real world.

Furthermore, it was suggested that the phenomena of interest was actually a somewhat permanent change in cognitive behavior and thus an outcome measure of retention rather than one of immediate learning was desirable.

However, in spite of these difficulties with much previous experimental work in the area of concept learning, it is true that much, if not most, of the empirical evidence that psychologists have collected about concept learning has been in the context of such experimental procedures. To make use of this wealth of experimental work requires that one use an experimental method which is similar enough to what has gone before to make use of that which others have learned.

The development of the specific set of concept learning tasks used in this study represent the simultaneous effort to satisfy several somewhat conflicting criteria. The tasks had to be reasonable analogies of the "learn the rule by which this collection of figures has been dichotomized" type of experiment. In addition they had to overcome the objections about the artificiality of these tasks which have been enumerated. Finally, the tasks or experimental treatments had to lend themselves to inclusion in the simplex design approach to a response surface model as has been described and to the study of the eight independent variables listed above

Since the need to overcome the artificiality present in most card-sorting or concept-board type of concept learning tasks was seen as the most important of these criteria, a decision was made to choose tasks which would consist of attempting to learn things which are typically taught in schools and about which there would be little disagreement as to their being correctly termed concepts. A major difficulty with this approach is the problem of getting a large supply of naive subjects--subjects who would enter the experimental situation knowing nothing of the concepts to be presented. One method of handling this difficulty would be to present concepts usually taught at one grade level to children at some

earlier grade level. This, in turn, requires that the concepts be ones that younger children are developmentally ready to handle and that are primarily school-taught concepts rather than the type the child could possibly have been introduced to via radio, television, newspapers, motion pictures or adult conversation.

The requirement that the tasks sufficiently resemble the typical card-sorting or concept-board type of task so that analogies could be drawn between the two meant that the set of concepts to be learned had to be such that it would be possible to specify the relevant attributes and the values of said attributes which define the concepts. Moreover, as one varies the attributes or values of the attributes, it should be true that one is creating exemplars of different concepts.

The third criterion is that the materials used had to fit the simplex design being studied and had to adapt to the specific set of variables chosen for inclusion in the design. Since this design calls for the determination of the $(k+2)^{nd}$ and succeeding experimental treatments only after the first (k+1) experiments have been run and since the total number of possible combinations that might be dictated by the basic set of combinations easily runs into the thousands,* the tasks

*Seven factors with five levels of each results in a complete factorial experiment with 78,125 experimental combinations.

chosen had to be such that a new experimental treatment combination could be put together on relatively short notice, in some cases overnight.

The first solution that suggested itself was the use of variations of colors, numbers and shapes of geometric forms that have been used in many concept experiments with the difference that these materials be used with children of a sufficiently young age that the children would indeed need to be learning a label for the concept and organizing their environment in novel ways while they were learning the concept. This, however, would have necessitated the use of pre-school or kindergarten age children and forced an individually administrated type task. This approach was eliminated on the grounds that it was not practical with the large numbers that seemed called for to provide sufficient precision in the estimation of responses. Another alternative considered was the use of a set of political concepts such as communism, socialism, democracy, dictatorship, etc. but this was rejected because of the difficulties in preparing a wide variety of exemplars of the concepts, in being able to identify precisely the defining attributes and in finding a population of students that was developmentally ready to work with concepts of this kind but which had not already had widely varying exposures to these concepts

via outside-of-school sources of information. Similarly, sets of language concepts and science concepts were rejected and the focus came back to mathematics, specifically geometry, in which concepts would lend themselves to visual presentations which could be carefully controlled.

The plane geometry concepts of trapezoid, isosocles triangle, quadrilateral, and rectangle were finally chosen to be learned by fifth-grade-age youngsters. These are concepts which have typically been taught in the sixth, seventh, and eighth grades which satisfies the criterion of being real concepts that schools attempt to teach. Furthermore, the newer, so-called modern math, approaches to teaching elementary school mathematics have pushed the first introduction of these concepts down into the third, fourth and fifth grades giving support to the contention that children of this age are ready, in a developmental sense, to acquire these concepts. In addition, these are concepts to which children do not get any appreciable exposure outside of school instruction. Finally, in the spring of 1965 there still remained numerous school systems in southwestern Wisconsin that had not made the transition to arithmetic series which incorporated the modern math approach and thus provided a large potential population of fourth and fifth graders for this study. Thus,

the criterion of being real concepts for which a large number of naive subjects could be found was satisfied.

A second criterion was that the relevant dimensions of the concepts be clearly specifiable and that varying the relevant attributes or the values of relevant attributes produce new and meaningful concepts so that close analogies could be made to the great bulk of concept learning studies which have been reported over the past fifteen years. In plane geometry we have a system which is quite rigorously and unambiguously defined. For example, all plane figures may be categorized as open or closed, all closed figures as polygons, non-polygons or composites of the two, and all polygons into a variety of categories or concepts depending upon the number of sides, the size of angles and the spatial relationships among sides and angles. Thus, within the context of a system of plane figures, one has a universe of concepts which, for practical purposes, is defined by a finite set of attributes or dimensions, each with a finite set of values. Yet, there exists an infinite number of perceptually different exemplars and non-exemplars of each concept. This infinity of exemplars is the result of the system including a number of irrelevant attributes, some of which may take on an infinity of values. Two of these are the size and spatial orientation dimensions.

While it makes sense to speak of an elementary school age child having a concept of a geometric figure such as a triangle, note that this is a relatively low level or unsophisticated concept of triangle compared to that of a high school geometry student or of a professional surveyor or of a mathematician specializing in geometry. Whether these differences should be explained as a series of steps in the complete learning of a concept or as a set of distinct concepts with a single label will not be argued here. It does, however, emphasize the problem of the relativeness and indeterminateness of that which educators would call concepts.

A third criterion had to do with the practical aspects of using these geometric concepts within the framework of the response surface exploration model of a sequence of experiments with the treatment combination at any point in the sequence being determined by responses to preceding experiments. By presenting the exemplars of the concepts by means of slide transparencies, a wide variety of treatment combinations was possible from a relatively small pool of photographic material by superimposing two or more transparencies in the same slide mount. Certain of the other variables included had to do with auditory cues and varying times which was controlled by a tape recorder which presented

verbal cues while regulating the pacing of slide presentation.

The concepts chosen to be used in the experiment were those of square and polygon, which were used as warm-up tasks, and quadrilateral, isosocles triangle, trapezoid, and rectangle which were presented as concepts to be learned and tested upon.

Of the eight variables designated to be varied during the set of experiments, seven were varied from experiment to experiment and the eighth, that relating to the relative complexity of the several concepts, within each experiment. At least five possible levels of the first seven factors were identified and three levels of the eighth. The factors and levels of each are presented in Table 2. Since all of the factors or variables were assumed to be at least ordinal in nature, they were coded numerically as -1, 0, 1, 2, 3. This choice of code was preferred to a 1, 2, 3, 4, 5 code because it allowed the original simplex to be oriented about the combination, (0,0,0,0,0,0,0), and permitted movement of the simplex design in all directions from the original set rather than only in a positive direction.

The actual choice of levels used is arbitrary and was based solely upon the judgment of the experimenter. Where empirical data is available it, of course, should

TABLE 2

Factors and levels of each included in the experimental design as independent variables.

Variable	Levels -1	0	1	2	3
1 Redundancy of information	3 instances	6 instances	9 instances	12 instances	15 instances
2 Mode of presentation	Singly	Singly: w/focus instance	With all previous instances	A sufficient set	An entire array
3 Ratio of positive to negative instances	All positive	2:1	1:1	1:2	Focus instance:all negative
4 Ordering of positive and negative instances	Randomly	All positive, then all negative	Alternating	Focus instance:all negative, then positive	
5 Verbal cue	Silence	This is a...	This is a... w/1 word hint	This is a... w/2 word hint	Statement of defining rule
6 Time projected	4 seconds	8 seconds	12 seconds	16 seconds	20 seconds
7 Time between slides	2 seconds	4 seconds	6 seconds	8 seconds	10 seconds
8 Complexity of concept	Quadrilateral (k+1)	Isosceles (k+2)	Trapezoid (k+2)	Rectangle (k+3)	

be given careful consideration in the making of these choices. In this situation, the existing empirical evidence was only tangentially related to the type of presentation being made. The decisions were based upon examination of the introduction of these concepts in several arithmetic textbook series, consultation with mathematics curriculum people, experience in analyzing data from concept attainment experiments performed at the Learning Laboratory of the Educational Psychology Department at the University of Wisconsin and consideration of the length of time one could hold the attention of a group of fifth grade children. Following the suggestions of Spendley <u>et al</u>. the levels were chosen so that a unit change on any one could reasonably be expected to result in a measureable shift in the response measure.

Thus, on the redundancy of information factor increments of three instances seemed adequate to produce measureable changes, if indeed this variable was important to the learning of the concepts. By using six instances as the zero level, the possibility of moving the simplex in a negative direction, i.e., to presentation of sets of three, two, or one instance was allowed.

The second factor is mode of presentation. Five modes that have been used in concept learning studies

formed the five levels of this factor. Theoretically, these five modes convey varying amounts of information to the subject and on this basis they may be ordered. That mode which contains the least information is the presentation of instances one at a time with no opportunity to refer to instances previously presented. A more informative mode of presentation is to identify the first positive instance of the concept as a focus instance and then present new instances in conjunction with this focus instance, thus providing a referent for the new instances. A further step is to present each successive instance in conjunction with all previously presented instances. A fourth level is to present a sufficient set of instances for the learning of the concept and to identify successive instances within the context of this set. A fifth level is the presentation of instances within the context of a much larger set than that identified as a sufficient set. In the laboratory type problem with a concept board this array is typically the entire universe of allowable instances. With a real world concept it is not possible to identify either a sufficient set or a complete set of exemplars of any one concept. To approximate the fourth mode of presentation a sufficient set of instances needed to define each concept, given that the instances were

plane figures, was devised and a larger display consisting of several of these pseudo-sufficient sets was used for level five. Sketches of the sets of figures used for the concepts trapezoid, isoceles triangle, quadrilateral and rectangle are presented in Appendix A.

The third factor is that of the ratio of positive to negative instances of the concept being learned. The five levels chosen were those of all positive instances, all negative with the exception of a single positive focus instance and three intermediate ratios of positive to negative, 2:1, 1:1, and 1:2. These levels were ordered on the basis of number of positive instances.

The fourth independent variable had only four levels. Four orders in which instances can be presented are to present the positive instances first, followed by the negative ones, the negative first followed by the positive or an interspersed arrangement where the order is either random or deliberate alternating of positive and negative. The ordering of these four levels was done by examining them in sets of three and asking which seemed to be an intermediate between the other two in terms of expected response. The difficulties with this ordering became apparent early in the experimental part of the study and the factor was eliminated in the contraction of the design after the basic set of experimental conditions were run.

The five levels of verbal cue to be used consisted of silence, the statement "This is (is not) a ..." followed by the appropriate name of the concept, the above statement with a one-word cue ("count" for the quadrilateral, "equal" for the isosceles triangle, "opposite" for the trapezoid and "corners" for the rectangle), the above statement followed by a stronger cue ("count sides" for the quadrilateral, "equal sides" for the trapezoid and "square corners" for the rectangle), and a complete definition of the concept.

The two time factors, time each slide was projected, and time between slide presentation of successive instances, were measured in seconds with slides projected for multiples of four seconds and the time between slides being multiples of two seconds. Consideration was given to the idea that times used might more closely represent an equal interval scale in terms of psychological impact if the levels were chosen as some exponential or geometric progression rather than an arithmetic one but this was rejected because it would have led to longer experimental sessions with the subjects than seemed practical for children of their age.

The eighth variable studied, the relative complexity of the concept, differs from the others in that it is not related to presentation of the task as are the first

seven. Instead it represents a change in task. Variation of this factor could have been carried out between experimental treatments by using only one concept per experiment or by using several of the same theoretical complexity, but it was believed that more could be learned about this factor by varying it within each experiment. As this meant multiple responses to any one experimental treatment, one level was chosen as the level of complexity to be used as the criterion response for the attempt to optimize on the other factors.

Since all the concepts used in this study are special cases of the concept, polygon, it is sufficient to specify the complexity of polygon as k and then specify the complexity of the several concepts used relative to k. Thus, a triangle is a polygon with a single added restriction as to the number of sides and therefore may be expressed as having complexity (k+1). An isosceles triangle is a triangle with the restriction that two sides (or two angles) be equal and therefore is of complexity (k+2). Similarly a quadrilateral is a four-sided polygon, complexity (k+1); a trapezoid is a quadrilateral with two parallel sides, complexity (k+2); and a rectangle is a trapezoid with right angled corners, complexity (k+3). Thus, if the complexity of a concept is important to the learning of that concept, quadrilateral should be easier to learn than either isosceles

triangle or trapezoid and all three should be easier to learn than rectangle.

The Experimental Procedure

The actual experimental treatment administered to any given group of subjects consisted of a short twenty to forty minute presentation. This presentation began with an informal statement by the experimenter* in which he introduced himself, briefly explained the connection with the University of Wisconsin, made some general comments about the weather, school, etc., to relax the children, and explained, generally, the purpose of the experiment was that of trying to find out some things about the ways in which youngsters learned. A short explanation of the synchronized slide and tape recording equipment was given to satisfy the children's curiosity as to how the machinery operated and thus have greater attention given to the actual presentation. With a final exhortation to pay close attention to the slides and tapes so as to learn as much as possible, the lights were turned off and the treatment presented. This introductory session was deliberately kept informal and somewhat uncontrolled in an attempt to get the youngsters into a relaxed frame of mind.

---

*All experiments were conducted by the author or Mary Davis of the staff of the Center for Learning and Re-Education at the University of Wisconsin.

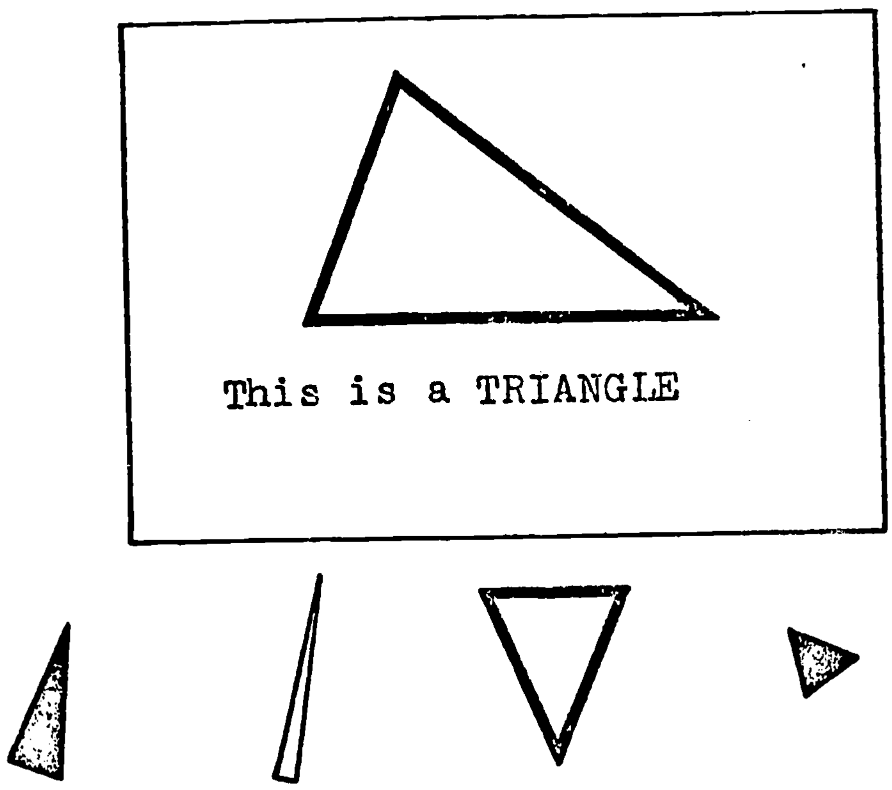


FIGURE 2

Focus instance slide for concept, triangle, and the succession of instances used to demonstrate the concept.

The slide and tape presentation (the complete transcript appears in Appendix B) began with a set of instructions specifying that the purpose of the task being presented was to discover what there was about certain types of geometric figures which resulted in their having a common name. This was followed by a practice set of five slides, all of which were positive exemplars of the concept triangle. Figure 2 shows the focus instance (the first positive exemplar) as it appeared on the screen and sketches of the other four exemplars used. Note that with triangle, as with all sets of instances, the attempt was made to present as wide as possible variety of triangles. Obtuse and

acute triangles, scalene, isosceles and equilateral triangles, triangles of various sizes and spatial orientation, and triangles as either solid or outline figures with lines of differeing widths were used. With each slide presented, the title, "This is a TRIANGLE" was simultaneously read via the tape recording. At the end of the sequence the children were asked if they now knew what made a figure a triangle. After a pause they were told that a triangle was any three sided figure.

A second practice set designed to elicit the concept, Square, was then presented. This practice set differed from the first in that it contained negative as well as positive instances of the concept. An example of one of the negative instances as well as the complete set of nine slides is presented in Figure 3. Again the youngsters were asked if they now understood what a square was and were then told that it was a figure with sides of equal length and square corners. It should be noted that toward the end of this practice sequence the children tended to become restless over the seeming triviality of the task. However, the appearance of the rectangle (which was not too different from a square) as a negative instance apparently forced many to focus more sharply on the materials being presented and emphasized the non-triviality of the task.

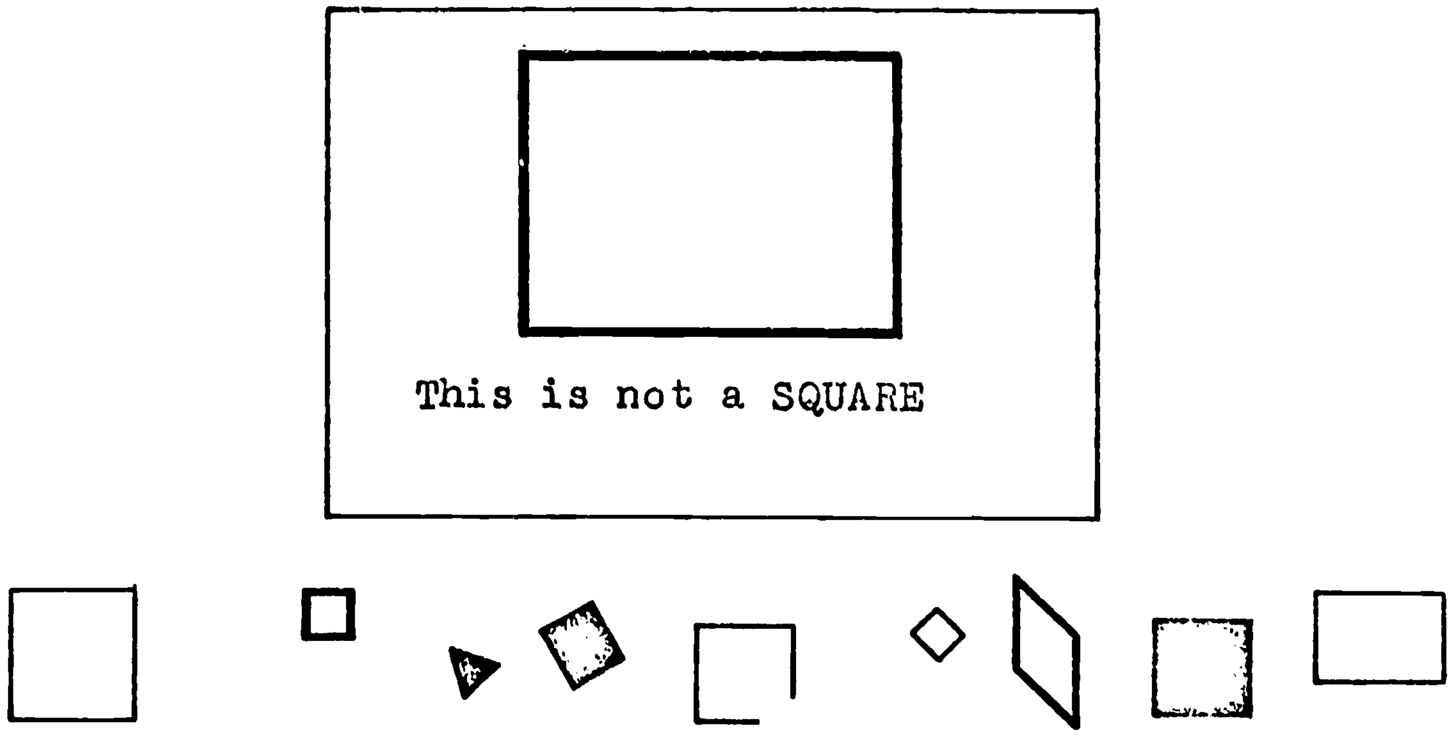


FIGURE 3

Negative instance of concept, square, and the series of instances used to demonstrate the concept.

The children were then told that the following figures would be much more difficult and then, in this order, were presented with sets of slides exemplifying the concepts of quadrilateral, isosceles triangle, trapezoid, and rectangle. These sets were systematically varied on the independent variables previously described. After each set the lights were turned on for one minute and the students were asked to perform an activity designed to reinforce what they had learned from the presentation. After the quadrilateral and trapezoid sets they were asked to try to write a definition of what they thought a quadrilateral or trapezoid was and after the

other two they were requested to try to draw two different examples of isosceles triangle or rectangle.

At the completion of the rectangle concept set the youngsters were told that the experimenter would be back the following day to see how much they had learned from watching the slides. It was suggested that it was what had been learned from the slides that was of interest, not what they could learn from their teacher, each other, their parents, books, or the other students who had not seen the slides. In the light of this interest they were requested to not talk about the experiment until after the experimenter returned, at which time they would receive a more complete explanation of what the experimenter was trying to accomplish. The following day, at approximately the same time of day, a return visit was made to the school and the response measure, a sixty-item test, was administered to both the control and experimental groups. Following the test, the experimenter defined the four concepts to the youngsters, explained how one could use the positive and negative instances to determine the definition and answered questions about the procedure.

The Response Measure

In the concept-board or card-sorting type concept learning experiment, the response measure has typically

1. Circle each figure below which is a QUADRILATERAL.
Do nothing to the other figures.

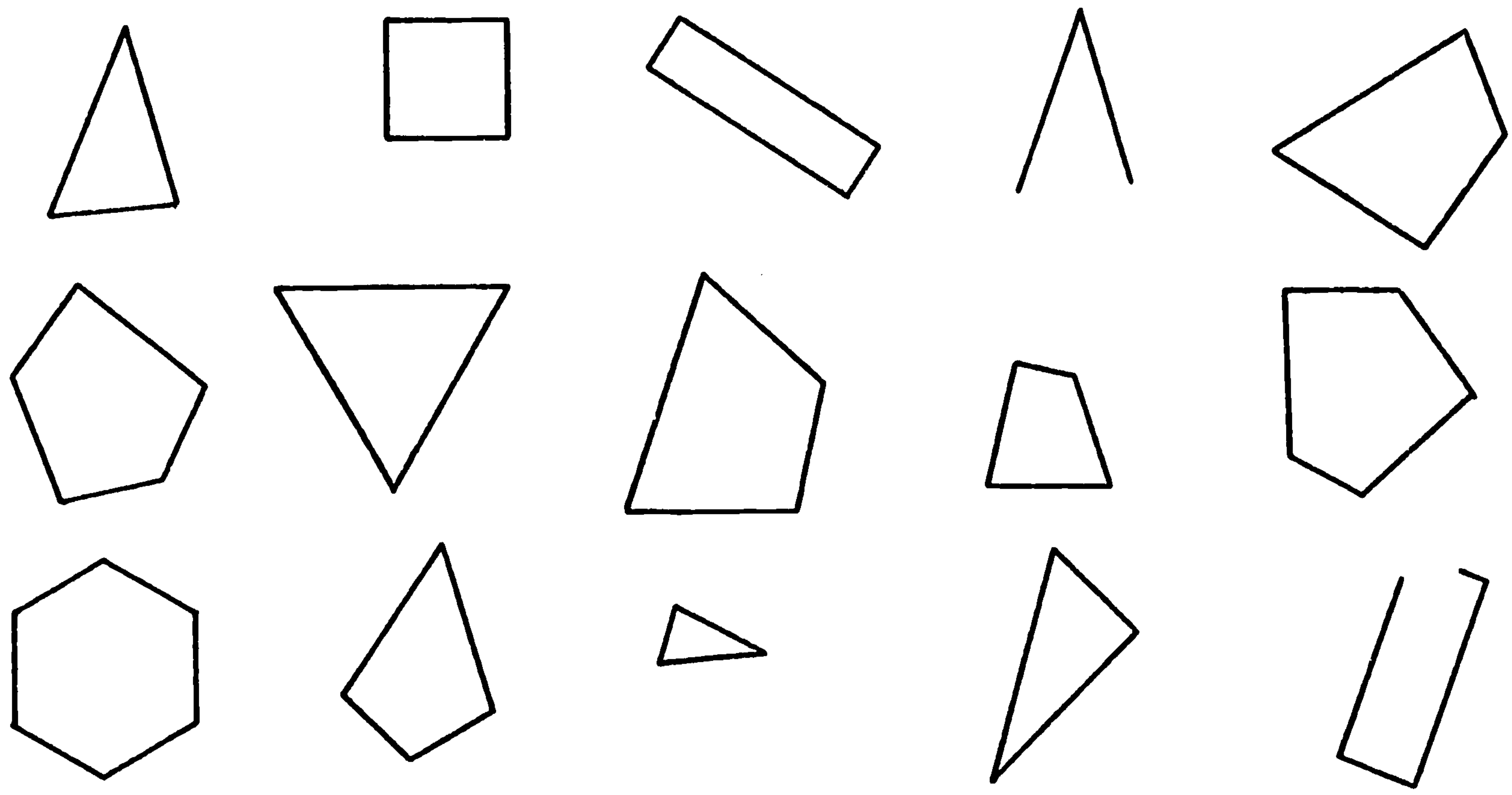


FIGURE 4

Item number 1 of the Geometric Figure Identification Test used to test for the concept, quadrilateral.

been some concomitant of immediate learning. The subject is expected to verbalize the rule which defines the concept, i.e., divides instances into exemplars and non-exemplars of the concept, or to do some task, such as sort cards, from which it may be inferred that the subject has acquired the rule. This dichotomy of responses is often converted into scores by considering some concomitant of success or failure, such as the number of trials needed, the amount of potential information received, the time required, or the number of

errors made. Typically, no attempt has been made to measure retention of the concept. This, of course, is not surprising since the artificiality of these tasks do not lend themselves to such a procedure.

In this set of experiments immediate learning was not measured nor were any concomitants of that learning. The criterion test, the Geometric Figure Identification Test, was given approximately twenty-four hours after the running of the experiment. This test, which appears in Appendix C, contained four items. Each of these items consisted of a set of fifteen geometric figures, some of which exemplified the concept being tested. The directions were to circle those figures which were exemplars of the concept. Figure 4 is the item for testing the concept, quadrilateral.

The Experimental Population

Several criteria had to be satisfied in choosing the population from which experimental units were to be chosen. The design called for a sequence of experiments, between fifteen and thirty in number, requiring a large supply of relatively homogeneous experimental material. As individual children are quite heterogeneous with respect to cognitive abilities, the use of groups of children rather than individual children as the units was indicated. A tryout of the criterion

test revealed that groups of approximately twenty children would reduce the standard deviation among group means to approximately one-half unit on the response measure. This implied the need for either a pool of 300 to 600 subjects from which groups could be randomly drawn or a pool of 15 to 30 schools which could be randomly assigned to treatments.

Another restriction was placed on the choice of population by the experimental task. Subjects or schools were needed which would provide naive subjects as regards the four geometric concepts being used. Specifically, this eliminated school systems which had adopted arithmetic tests which introduced these concepts in the middle and upper elementary grades.

A third criterion that had to be satisfied was the practical one of having the students in a sufficiently compact geographical area to permit the running of the experiments with reasonable costs.

The two large metropolitan school systems within feasible geographic limits were both rejected because both were using arithmetic text series which introduced the concepts being used before the sixth grade. Had this not been the case, the alternative of running the experiment within a single large system would still have been rejected. The transporting of children necessary

for the random choice of subjects from a single large pool of subjects was deemed impractical and the hetrogeneity among schools within a single system was thought undesirable.

The alternative chosen was to identify a large number of school districts representing similar communities, choose those that fit the criteria and then randomly choose schools to provide the experimental group. The specific criteria used follows:

1. Neither the arithmetic series being used nor the experience of the state mathematics supervisor indicated that fifth grade children would have had any instruction on the concepts being studied.
2. The school was the only elementary school serving the town and surrounding area in which the school was located.
3. The school was large enough to have at least two rooms (approximately 40 children) of fifth grade students but no more than three.
4. The school was located within ninety minutes' driving time of the University of Wisconsin.

The first of these criteria aimed at providing naive subjects. The second provided some assurance that the school populations would be relatively similar from school to school since each school would serve an entire

community. The third criterion made it possible to select an experimental group of twenty children and still have an equal number to use as a control group so that the similarity of the several schools could be tested. The upper limit on the size of the fifth grade reduced the possibility of drawing a random sample that was unrepresentative of that fifth grade, as well as setting limits which guaranteed communities within a specific size range. The fourth criterion was a practical necessity in carrying out the logistics of the experiment.

Thirty schools were identified that met the above criteria.* These schools were then put in random order. Letters (see Appendix D) were sent to the administrators of the first twenty schools on the random list eliciting their participation in the project. Eighteen of the twenty responded favorably, and a nineteenth indicated that he felt pressured to refuse only due to teacher problems at the fifth grade level. The twentieth indicated that he thought research of this type an unwarranted imposition on public schools. As time only permitted the completion of eighteen experiments the remaining ten schools were never contacted. A complete list of the cooperating school systems is found in

*Two exceptions were made to criterion 2 in order to obtain a sample of 30. These were in two 2-school communities where the differences between the schools and the area served by the schools were judged minimal.

Appendix E.

This chapter presented the details and rationale for the development of the experimental design, the choice of variables to be studied and specifics about the levels of the several variables, the selection of the experimental task, the development of the experimental treatment, and the choice of subjects. While this chapter has presented these topics as discrete entities, this is not how they are developed. Rather, the many decisions to be made were formed through the simultaneous refinement of all these various aspects.

CHAPTER IV

RESULTS

This chapter, which contains the analysis and summary of the data from the sequence of experiments, is divided into two parts. The first examines the experimental data as it applies to the purpose of empirically testing the use of the simplex design in attempting to optimize responses in an educational experiment. The second examines the data as it relates to the previous body of knowledge in the area of concept learning.

The set of independent variables studied and the codes used to identify the levels of each are described in Table 2 on page 73. The sequence of sixteen experiments that were performed are numbered from 1 to 15, with the subscript a indicating a replication or re-run due to difficulties connected with the first running of that combination.

The Simplex Design in Optimizing Responses

The aim of each experiment was to present four concepts, of three different complexities, to be learned by a group of fifth grade children. Amount of learning was measured by a Geometric Figure Iden-

tification Test in which each child was presented with four sets of fifteen figures from which the child was to correctly identify the figures which exemplified the concept. Thus, each child received four scores which varied from 0 to 15, one for each of the concepts, quadrilateral, isosceles triangle, trapezoid, and rectangle. As has been described in Chapter III, these concepts are of three different theoretical degrees of complexity, the first of complexity (k+1), the second and third (k+2) and the fourth (k+3). For the purpose of studying the effectiveness of the simplex model in optimizing responses, the total score on the two (k+2) degree concepts, isosceles triangle and trapezoid, was taken as the criterion response. As there was no assurance that previous arithmetic instruction or ability would be the same from one set of subjects to the next, each group was randomly divided into an experimental and a control group and the mean score of the experimental group was adjusted by subtracting from it the mean score of the corresponding control group. This measure, the difference in mean scores between experimental and control groups on the isosceles triangle and trapezoid items, was the response used in attempting to arrive at an optimal combination.

The basic simplex design consisted of the eight combinations presented in Table 3. Two Number 5

combinations are listed because that particular experiment was replicated when a teacher made some remarks at the conclusion of the experimental treatment which might have influenced the control group. As the results of the replication indicated that these fears were unfounded, the first Number 5 combination was used in the analysis. The numbers in the several experimental and

TABLE 3

The set of eight treatment combinations defining the basic 7-dimensional simplex.

	Variables*	1	2	3	4	5	5a**	6	7	8
1	Redundancy of info.	0	1	0	0	0	0	0	0	0
2	Type of presentation	0	0	1	0	0	0	0	0	0
3	Ratio of + to -	0	0	0	1	0	0	0	0	0
4	Order of + and -	0	0	0	0	1	1	0	0	0
5	Verbal cue	0	0	0	0	0	0	1	0	0
6	Time per slide	0	0	0	0	0	0	0	1	0
7	Time between slides	0	0	0	0	0	0	0	0	1
N in Experimental group		24	24	21	24	21	21	20	26	24
N in Control group		43	32	33	65	35	77	25	35	79

* Complete code for variables and levels are found in Table 2 on page 73.

** Experiment 5 was replicated.

control groups are also presented in Table 3. While the original decision was to have twenty children in each experimental group, these numbers are between 20 and 26 as additional experimental subjects were chosen to guarantee that absences on the test day would not result in less than 20 in a group.

The results of these nine experiments are presented in Table 4. The mean scores for the experimental groups were in the range, 19.33 to 22.85. Those combinations which contained higher levels of verbal cue

TABLE 4

Summary statistics for the basic set of experimental combinations on the combined score of the isosceles triangle and trapezoid items.

Experimental Combination Number	$\bar{E}$	$\bar{C}$	s_{pooled}	$(\bar{E}-\bar{C})$*	$s_{\bar{E}-\bar{C}}$
1	21.04	16.21	3.63	4.83	.92
2	21.58	16.72	3.42	4.86	.92
3	20.90	15.97	3.53	4.93	.99
4	20.50	14.99	3.00	5.51	.72
5	19.33	15.60	3.19	3.73	.88
5a	19.71	15.62	3.63	4.09	.89
6	22.85	16.40	3.51	6.45	1.05
7	21.46	14.23	3.48	7.23	.90
8	19.66	15.84	3.27	3.82	.76

*All differences are significant at $p<.001$.

(Number 6), redundancy of information (Number 2), and time per slide (Number 7) produced the largest absolute responses while those which were set at higher levels of order of presentation (Number 5a and 5b) and increased time between slides (Number 8) resulted in the smallest responses. This was consistent with the results in terms of the corrected response, $\bar{E}-\bar{C}$.

The mean scores of the control groups are in the range, 14.23 to 16.72. The expected score, under the

assumption of no knowledge and independent guesses on each item, would be 15.00. Every control group had one or more members with scores of 23 to 26 suggesting at least partial knowledge of the concepts. This is to be expected, particularly on the isosceles triangle concept to which children could react solely to the word "triangle." The result is mean control group scores slightly above chance levels.

Examination of the standard deviations of the experimental and control groups revealed no consistent pattern that would indicate that the experimental treatments were operating to either increase or decrease the variability among subjects. Therefore, a pooled estimate of the standard deviation was calculated for each experiment. These pooled standard deviations were quite consistent, all being between 3.00 and 3.63.

The differences between mean experimental and control scores ranged from a low of 3.73 for combination 5 to a high of 7.23 on combination 7. The standard errors of the differences in mean scores all fell between .72 and 1.05, and all the obtained differences between experimental and control group means were at least 4.23 standard errors in size. Translating this into tests of null hypotheses of no difference between experimental and control group means, all the differences

between experimental and control groups are significant ($p < .001$).

Of this basic set of eight combinations, the minimal response in terms of difference between means of experimental and control groups was from experimental combination 5, a difference of 3.73 points. Thus, the decision was made to reject this combination for a new one.

Following the rule suggested earlier, i.e., subtract the coordinates of the discarded point from twice the mean of those from the remaining points, resulted in the set of values ($\frac{1}{4}$, $\frac{1}{4}$, $\frac{1}{4}$, -1, $\frac{1}{4}$, $\frac{1}{4}$, $\frac{1}{4}$). Rounding these values off to the nearest integer values gives combination Number 9: (0,0,0,-1,0,0,0). The experiment run at this combination produced a mean of 21.37 for the Experimental group and a mean 15.78 for the control group. This led to a difference between the means, ($\bar{E}-\bar{C}$), of 5.59 which, with a standard error of 1.04, was significant with $p < .001$. This mean difference was greater than that obtained for combinations 1-4 and 8. Among the responses from 1-4 and 6-9, that from combination 8 was low and should have been the next combination replaced.

At this point pressures of time forced a reduction in the complexity of the experimental treatments.

Delays in getting additional photographic materials processed, the re-running of combination 5, and postponements requested by schools had cost over two weeks of time which could not be made up because the experimental period was limited by the closing of schools in late May. A decision was made that more could be learned in the period of time remaining by reducing the number of variables from seven to four. On the basis of their intrinsic interest to the experimenter as well as on the results of the ten experiments which had been completed, the variables of ratio of positive to negative instances, order of positive and negative instances, and time between slides were fixed at the 0 levels of each. Combinations 4, 5, and 8 were removed from the original design leaving combinations 1, 2, 3, 6, and 7 to define a four-dimensional simplex. This reduction of the design from seven to four dimensions and its effect on succeeding experiments is given in Table 5.

Of these five combinations, Number 1 had the low response. Recalculating the next experiment resulted in the combination ($\frac{1}{2}$,$\frac{1}{2}$,$\frac{1}{2}$,$\frac{1}{2}$) which rounded off to (1,1,1,1) and is Number 10 in Table 5. A sequence of five additional experiments were run and the design of each is given in Table 5. Combination 11 replaced 2;

TABLE 5

The complete sequence of experiments as combinations of the seven independent variables and the final reduced simplex sequence as combinations of four variables.

Variable	Experimental Combination															
	1	2	3	4	5	5a	6	7	8	9	10	11	12	13	14	15
Initial Simplex																
1 Redundancy	0	1	0	0	0	0	0	0	0	0	1	0	1	0	1	0
2 Type of presentation	0	0	1	0	0	0	0	0	0	0	1	1	0	1	0	1
3 Ratio: + to -	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
4 Order	0	0	0	0	1	1	0	0	0	-1	0	0	0	0	0	0
5 Verbal cue	0	0	0	0	0	0	1	0	0	0	1	1	0	1	1	0
6 Time/slide	0	0	0	0	0	0	0	1	0	0	1	1	0	1	1	0
7 Time between slides	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Final Simplex																
1 Redundancy	0	1	0				0	0			1	0	1	0	1	0
2 Type of presentation	0	0	1				0	0			1	1	0	1	0	1
5 Verbal cue	0	0	0				1	0			1	1	0	1	1	0
6 Time/slide	0	0	0				0	1			1	1	0	1	1	0
Replaced											1	2	11	12	3	14
Replaced by	10	11	14				*	*			*	12	13	*	15	*

* Final simplex consisted of combinations 6. 7, 10, 13, and 15.

12 replaced 11; 13 replaced 12; 14 replaced 3; and 15 replaced 14. If time had been available to run a longer sequence, 15 would have been the next combination replaced. The final simplex consisted of the combinations 6, 7, 10, 13, and 15.

The statistics from these additional combinations are given in Table 6. The pooled standard deviations for combinations 9 through 15 are of the same approximate size as those from the basic set of combinations as are the standard errors of the difference in means.

The response to Number 11 demonstrates the usefulness of the simplex design in the presence of experimental error. The difference in mean scores, $(\bar{E}-\bar{C})$, is .78, the only difference which is not significant ($p<.01$). In talking to the teachers after this experiment was completed, it was revealed that the teachers in this school had spent several days teaching these concepts earlier in the semester. If this had not been revealed by the teachers, it would have been obvious from the analysis of the test results. While this evidence was sufficient to justify discarding that experiment and replicating it on another group of students, this decision was not made. Instead the result was allowed to stand; the idea being to observe to what results the decision rules would lead. The response to Number 11 being low it was replaced by

TABLE 6

Summary statistics* on the isosceles triangle and trapezoid items for all experimental combinations.

Experimental Combination	$\bar{E}$	$\bar{C}$	$\bar{E}-\bar{C}$	s_{pooled}	$s_{\bar{E}-\bar{C}}$
1	21.04	16.21	4.83	3.63	.92
2	21.58	16.72	4.86	3.42	.92
3	20.90	15.97	4.93	3.53	.99
4	20.50	14.99	5.51	3.00	.72
5	19.33	15.60	3.73	3.19	.88
5a	19.71	15.62	4.09	3.63	.89
6	22.85	16.40	6.45	3.51	1.05
7	21.46	14.23	7.23	3.48	.90
8	19.66	15.84	3.82	3.27	.76
9	21.37	15.78	5.59	3.26	1.04
10	20.66	13.15	7.16	3.63	.95
11	21.00	20.22	.78	3.59	1.12
12	21.30	16.46	4.84	3.25	.97
13	22.18	17.05	5.13	3.78	1.01
14	20.73	17.26	3.47	3.88	1.16
15	19.71	15.79	3.92	2.85	.90

* Maximum score was 30.

Number 12 which, it should be observed, is the same as Number 2, which 11 had replaced. The difference in mean scores, $(\bar{E}-\bar{C})$, was 4.84 which was again low and led to Number 13 which is identical to combination Number 11. Thus, our decision rules led to the same end result as common sense, i.e., combination Number 11 should be performed on a new group of subjects. This time the combination produced a response, $(\bar{E}-\bar{C})$, of 5.13 which was not low.

The final two experiments consisted of replacing 3 with 14, and, when that proved low, replacing 14 with 15 which was identical to 3. Fifteen also was low and would have been replaced had there been time to continue the sequence. In calculating combination 14 the values (½,0,1½,1½) were rounded off to (1,0,1,1) instead of (1,0,2,2) because the latter value, in the judgment of the experimenter, resulted in too large an increment with corresponding distortion of the shape of the simplex. That experiments 14 and 15 led to this same set of calculated values indicates that experimenter's judgment was probably wrong and the 16th combination would have been (1,0,2,2).

The five combinations, 1, 2, 3, 6, and 7, which formed the basic simplex in the reduced four-variable system had a mean response (mean of$(\bar{E}-\bar{C})$) of 5.66. The four combinations (omitting 15 which would have

been replaced) left in the final simplex had a mean response of 6.49. This represented an increase of 0.83 over the sequence of six additional experiments.

While the sequence of experiments was short, five in the basic set and six additional ones, several points should be noted. The mean response of the final simplex was nearly one standard error greater than the mean response of the initial simplex. Reducing the number of variables from seven to four at the completion of experiment 9, demonstrated the manner in which the design can be collapsed. That combinations 11 and 12 and 14 and 15 merely offset one another is evidence that the decision rules will work in the presence of experimental error. Since none of the eleven experiments in the reduced design involved more than two levels of any one factor, the advantage over a two-level factorial design is not readily apparent. However, when one considers that one of the experiments, combination 11, could have been eliminated on the basis of reports from the teachers involved and that the next experiment in the sequence would have been at a third level of two of the factors, the potential saving in experimental effort over a 2^4 or 3^4 factorial is more obvious.

The Results Applied to Concept Learning

As was detailed earlier eight variables were identified and built into the experimental design, in part because of their relevance to concept learning and in part because of the ease with which they could be manipulated in an empirical test of the simplex design model for optimizing responses. Sixteen experiments were performed with the levels of each of these variables determined by the criterion of obtaining the maximum possible response over a set of points held in a specific configuration relative to each other. This criterion does not lend itself to the simultaneous investigation of the tenability of specific hypotheses about the variables although it seemed reasonable that much useful information about the relationship between concept learning and the independent variables would be forthcoming as a by-product of the investigation of the simplex design.

Table 7 presents mean scores on all four concepts for both experimental and control groups for all sixteen experiments. The differences between these mean scores, standard errors for these differences, t-ratios and significance levels appear in Table 8. A general picture of the data in Table 7 can be obtained by studying the medians of the group means which are reported at the bottom of the table. In general, both experimental

TABLE 7

Mean scores for both experimental and control groups on four geometry concepts at eighteen experimental treatment combinations.

Experimental Combination	Quadrilateral $\bar{E}$	Quadrilateral $\bar{C}$	Isosceles Triangle $\bar{E}$	Isosceles Triangle $\bar{C}$	Trapezoid $\bar{E}$	Trapezoid $\bar{C}$	Rectangle $\bar{E}$	Rectangle $\bar{C}$
1	9.67	9.30	10.78	8.40	10.25	7.81	10.21	10.42
2	10.92	9.13	10.54	8.56	11.04	8.16	12.38	10.94
3	10.29	9.24	11.52	8.45	9.38	7.52	12.48	10.82
4	10.21	8.88	11.25	7.54	9.25	7.45	12.33	11.52
5	9.19	9.11	11.52	8.26	8.81	7.34	11.90	11.06
5a	9.52	8.77	10.00	8.04	9.71	7.58	9.52	10.22
6	12.00	9.80	12.15	8.84	10.70	7.56	12.45	12.00
7	9.81	8.46	11.23	7.06	10.23	7.17	12.54	11.57
8	8.46	8.54	10.42	7.95	9.08	7.89	11.47	10.54
9	11.42	9.06	10.37	7.89	11.00	7.89	12.95	11.84
10	11.42	8.53	10.58	6.92	10.08	6.58	12.21	10.53
11	10.75	9.44	11.37	10.72	9.63	9.50	12.13	11.72
12	11.09	8.54	11.04	8.23	10.26	8.23	11.74	10.86
13	10.77	10.13	11.36	9.82	10.82	7.23	11.68	11.74
14	10.55	8.74	10.82	8.83	9.91	8.43	12.27	9.39
15	10.00	8.53	10.19	7.95	9.52	7.84	11.43	11.42
Median	10.59	9.06	10.80	8.26	10.00	7.81	12.02	10.94

and control groups did best on the concept, Rectangle, with the mean scores of all 32 groups being above the chance expectation of 7.50, with the control groups in a range from 9.39 to 12.00 and the experimental groups from 9.52 to 12.95. On the concept, Quadrilateral, control groups also did considerably better than chance with mean scores in the range 8.53 to 10.13. Mean scores for the experimental groups were between 8.46 and 12.00.

The two concepts of medium complexity, Isosceles Triangle and Trapezoid, proved more difficult for both control and experimental groups with the controls averaging only slightly better than expected by chance with the median of the control group means being 7.81 for Trapezoid and 8.26 for Isosceles Triangle. These two concepts also produced the greatest differences between controls and experimentals. On the Trapezoid concept the control group means were between 6.58 and 9.50 while the experimentals were between 8.32 and 11.04. The corresponding ranges for Isosceles Triangle were 6.92 to 10.72 for the control groups and 10.00 to 12.15 for the experimental groups.

Table 8 presents this same general picture in terms of the significance tests of the differences between means of experimental and control groups. Examination of the table reveals that only four of the treatment

combinations succeeded in producing a significant increase in the learning of the concept, Rectangle, while all except two treatments produced significant increases for both the concepts, Isosceles Triangle and Trapezoid. On the concept, Quadrilateral, 9 of 16 combinations resulted in significantly higher scores by the experimental groups.

While the results of a short pilot study and an investigation of curricular materials had indicated that the four concepts chosen would be equally unknown to most fifth graders, this obviously did not turn out to be true. The scores on the Rectangle item indicate that many fifth graders have a partial concept of rectangle and that the experimental treatments did little to improve this concept.

Results of control groups on the Quadrilateral items also suggest some partial knowledge of that concept. Another explanation is that youngsters with an adequate concept of triangle could eliminate certain items on the test while guessing at the rest and, thereby, attain a better than chance score. In fact, one who knew what a triangle* was and eliminated the three

*An examination of a sample of 101 tests by control subjects showed that 79 correctly identified all three triangles as not-quadrilaterals.

TABLE 8

Differences between mean scores, standard errors of the differences, t-values and significance levels for each of the four concepts and fifteen experimental treatment combinations.

Experiment	Quadrilateral $\bar{E}-\bar{C}$	Quadrilateral $s_{\bar{E}-\bar{C}}$	Quadrilateral t	Isosceles Triangle $\bar{E}-\bar{C}$	Isosceles Triangle $s_{\bar{E}-\bar{C}}$	Isosceles Triangle t	Trapezoid $\bar{E}-\bar{C}$	Trapezoid $s_{\bar{E}-\bar{C}}$	Trapezoid t	Rectangle $\bar{E}-\bar{C}$	Rectangle $s_{\bar{E}-\bar{C}}$	Rectangle t
1	.37	.57	.65	2.38	.64	3.71***	2.44	.58	4.18***	- .21	.92	- .22
2	1.78	.70	2.54*	1.98	.68	2.93**	2.88	.66	4.34***	1.44	.75	1.92
3	1.05	.74	1.42	3.07	.80	3.85***	1.86	.54	3.44**	1.66	.60	2.78**
4	1.33	.65	2.04*	3.71	.53	7.07***	1.80	.50	3.60***	.81	.43	1.88
5	.08	.52	.15	2.26	.70	3.22**	1.47	.60	2.44*	.84	.78	1.07
5a	.75	.51	1.47	1.96	.66	2.97**	2.13	.54	3.92***	- .70	.74	- .95
6	2.20	.73	3.03**	3.31	.75	4.42***	3.14	.64	4.92***	.45	.79	.57
7	1.35	.66	2.04*	4.17	.54	7.69***	3.06	.60	5.08***	.97	.80	1.22
8	- .08	.53	- .15	2.47	.53	4.70***	1.19	.50	2.38*	.93	.45	2.06*
9	2.36	.73	3.26**	2.48	.85	2.91**	3.11	.66	4.73***	1.11	.91	1.21
10	2.89	.65	4.48***	3.66	.70	5.27***	3.50	.56	6.07***	1.68	.54	3.09**
11	1.31	.94	1.40	.65	.77	.85	.13	.73	.18	.41	.59	.69
12	2.55	.86	2.97**	2.81	.63	4.50***	2.03	.76	2.69*	.88	.97	.91
13	.64	.74	.86	1.54	.73	2.11*	3.59	.54	6.70***	- .06	.46	- .13
14	1.81	.90	2.02*	1.99	.79	2.53*	1.48	.74	2.01	2.88	.80	3.61***
15	1.47	.71	2.08*	2.24	.69	3.30**	1.68	168	2.48**	.01	.84	.01

* Significance level = .05
** Significance level = .01
*** Significance level = .001

triangles as possible quadrilaterals while guessing at the remaining twelve items would have an expected score of 9.00 which is quite close to the 9.06 median of control group means. Examination of the t-ratios, in the light of this interpretation, leads to the conclusion that there was less learning of the Quadrilateral concept than of the Isosceles Triangle or Trapezoid ones but considerably more than of Rectangle.

The Isosceles Triangle and Trapezoid concepts consistently resulted in significantly greater achievement by the experimental groups with only Number 11, for which the results were influenced by prior instruction, and Number 14, for which the t-ratio on the trapezoid concept was 2.01 as opposed to a critical value at the .05 level of 2.02, not producing significant differences. Control group means for the Isosceles Triangle concept were slightly above chance values (median value was 8.26), probably for similar reasons to those discussed above. Although all except one of the fifteen items making up the question on Isosceles Triangle were perceptually similar to, or actually were, triangles, knowledge of the concept of triangle could result in eliminations which would alter the chance expectations. On the Trapezoid items, with almost no extraneous cues to aid the subject (no triangles were used), the median of the control group means was 7.81, only .31 above chance expectation.

The responses to the treatments varied among concepts and general statements about the efficacy of the various treatments across all concepts are difficult to formulate. One approach is to rank order the experiments and examine those treatment combinations which uniformly have high ranks. Since there seems to have been partial knowledge of the Rectangle concept apart from that gained through the several treatments, only the other three will be considered. Table 9 presents the treatment combinations, their rank ordering on each of the three concepts (in terms of $\bar{E}$-$\bar{C}$) and the mean ranking of each treatment. Studying the mean rankings indicates that combinations 10, 7, 6, 4, 9, and 12 were the most effective over all concepts.

The result that Number 10 was most consistent in producing high levels of learning across all concepts is not unexpected as it represents the strongest combination presented. Nine instances were used instead of six, succeeding instances were presented against a background of all preceding instances, a verbal cue accompanied the slide presentation and the instnces were projected for 12 rather than 8 seconds.

One might have expected combinations 11, 13, and 14 to yield the next best results, as they contained the one level of three of the four variables in the final design. However, these three combinations yielded

TABLE 9

Treatment combinations, rank orderings of the differences in means ($\bar{E}-\bar{C}$) for three concepts and the means of these three rankings.

Comb. No.	Variables 1	2	3	4	5	6	7	Rank Order* on ($\bar{E}-\bar{C}$) Quad.	Isos. Tri.	Trap.	Mean Ranking
1	0	0	0	0	0	0	0	14	8	7	9.7
2	1	0	0	0	0	0	0	5	12	6	7.7
3	0	1	0	0	0	0	0	11	7	10	9.3
4	0	0	1	0	0	0	0	7	2	9	6.0
5	0	0	0	1	0	0	0	15	10	13	12.7
5a=5	0	0	0	1	0	0	0	10	11	8	9.7
6	0	0	0	0	1	0	0	3	6	4	4.3
7	0	0	0	0	0	1	0	7	1	3	3.7
8	0	0	0	0	0	0	1	15	4	14	11.0
9	0	0	0	-1	0	0	0	2	13	5	6.7
10	1	1	0	0	1	1	0	1	3	2	2.0
11	0	1	0	0	1	1	0	12	16	16	14.7
12=2	1	0	0	0	0	0	0	4	5	11	6.7
13=10	0	1	0	0	1	1	0	13	15	1	9.7
14	1	0	0	0	1	1	0	9	14	15	12.7
15=3	0	1	0	0	0	0	0	7	9	12	9.0

* Rank 1 indicates largest ($\bar{E}-\bar{C}$) difference.

consistently poor results, indicating an interaction between the variables of mode of presentation and amount of redundancy when combined with the added verbal cue and increased time per instance. The added redundancy with only the focus instance presented simultaneously and the presentation of a more limited set of instances against a background of all previous instances both led to poor results when combined with the 1 levels of verbal cue and time per slide. The response to these three combinations yielded results somewhat lower than those combinations which contained the high level of only one variable.

The basic set of eight experiments consisted of one at the low or 0 level of all seven variables and one each with a single variable set at the next higher or 1 level. Examination of the results of these experiments, either in terms of the difference between means of the two groups (Tables 6 and 8) or in terms of the mean of the experimental group (Table 7) shows that the variables of verbal cue (Number 6) and time per slide (Number 7) produce the greatest effect. Compared to the base combination, Number 1, the addition of a one word verbal cue increased the response ($\bar{E}-\bar{C}$) from 4.83 to 6.45. Increasing the time each instance was available from 8 seconds to 12 seconds resulted in an increased response of 7.23.

The only other single variable that produced an effect substantially above that of the base combination was that of ratio of positive and negative instances (Number 4). The change from all positive instances to a ratio of two positive to one negative led to an increased response of 5.51.

In summary, several points about the concept attainments aspects of the experimental sequence should be emphasized. Measurable, significant amounts of concept learning were obtained consistently from extremely short exposures to the sets of instances of each concept. This learning was not measured immediately, but after a delay of twenty-four hours and the amount of learning compared favorably to that of a group which inadvertently had had three arithmetic lessons on these concepts approximately ten weeks earlier. Examination of performances across the several concepts indicates no relationship between theoretical complexity of the concepts and the obtained responses. Finally, of the eight variables included in the experiments, those of verbal cue, time per slide, and ratio of positive to negative instances led to the largest increments in concept learning.

CHAPTER V

SUMMARY AND CONCLUSIONS

This dissertation has had two major foci, that of response surface methodology and that of concept learning. The experimental techniques and strategies of the RSM model, which have proved to be extremely powerful in a wide variety of industrial applications, have never been applied to a research problem in the behavioral sciences. While reasons for this lack of application may be easily identified, basic similarities between industrial and educational production do exist and seemed to justify the attempt to apply some of the RSM ideas to educational problems. The psychological area of concept learning is one in which much empirical work has been done--important variables have been identified and gross relationships among variables established. Furthermore, the learning of concepts would seem a major purpose of schools. Yet, applications of the findings of concept learning studies to experimental investigations of learning in schools are relatively difficult to find. It has been suggested that this paucity of applications may be due to difficulties with the experimental task used to simulate concept learning behavior.

Since the body of empirical evidence in the concept learning area has identified a number of variables which approximated the requirements of the RSM model, the major purpose of this dissertation has been the application of the exploration phase of RSM to a concept learning task in a school setting. The results of the investigation are best discussed separately under the headings, concept learning and response surface exploration.

Concept Learning

A sequence of sixteen experiments was run over a two-month period. While the sequence was not long enough to lead to an experimental combination producing a maximum response, there were some interesting byproducts of individually examining the results of the several experiments. These include:

1) Fifteen of sixteen experimental groups did significantly better than the corresponding control group on the isosceles triangle and trapezoid items of the response measure. It should be noted that these results were obtained from extremely short exposures (between 48 and 108 seconds per concept) to the relevant sets of instances. The sole exception was one group that had received prior instruction

on these concepts in the course of arithmetic lessons.

2) Lack of significant differences between experimental and control groups in the schools which had taught a unit on geometry some months earlier was strictly a function of higher scores by the control group. On none of the four concepts was the experimental group's mean response high and on the trapezoid concept it was less than the median of the other experimental groups.

3) The single variables that elicited the greatest improvement when presented at higher levels were those of verbal cue and time per slide. The improvement due to an increase in the proportion of negative instances used was the only other result substantially higher than that of the base combination (the 0 levels of all variables).

4) The complexity of the concept to be learned appeared to have no systematic effect. The two medium complexity concepts, Trapezoid and Isosceles Triangle, elicited the largest differences between experimental and control group means while the high complexity concept, Rectangle, produced the smallest difference. The uniformly high performance of control groups

on the rectangle concept points to prior knowledge of the concept indicating that it was probably a poor choice.

From the above points several generalizations can be drawn.

It is possible to obtain measurable amounts of concept learning by extremely short exposure to a series of positive and negative instances of that concept and without resorting to a definition of the concept. The variability in the effects of unit changes in the seven variables in the basic design does not necessarily reflect relative importance or strength of the variables as it could be a function of the units chosen by the experimenter. However, the addition of a single word, attention-directing verbal cue, and the increase of four seconds in the time the subjects had to study each instance did produce responses substantially above the base levels of each, while an increase in the number of instances from 6 to 9 and presenting each new instance along with all previous ones did not.

That verbal cue and the ratio of positive to negative instances were two of the three variables producing substantial increments above the base level may be a function of the indeterminateness of the universe within which the instances were being presented. That

is, because the universe was ill-defined, those changes in the treatments which directed attention to relevant dimensions or narrowed the frame of reference resulted in the largest increments in learning.

An examination of the geometry materials in fifth grade arithmetic texts will show that concepts such as those used here are presented almost entirely with positive instances, often only one, and a written definition. The fortuitous, but accidental, inclusion of a school in which the concepts had been taught with this general approach provides some insight into the effectiveness of that instruction. One might expect that the experimental treatment would have operated to strongly reinforce the earlier learning with resultant test scores considerably higher than other experimental groups. This did not happen, suggesting that classroom instruction might benefit from more emphasis upon the unique aspects of the approach used here, specifically the inclusion of more negative instances of the concept, a wider variety of instances, and allowing the student to infer the concept witnout a definition.

Finally, the idea of the complexity of a concept, which is well defined in a finite universe, may not be the same in an infinite universe. There was no consistent order across the complexity dimensions but this may have

been a result of the stated difficulties with the rectangle concept.

Response Surface Exploration

The attempt to explore the response surface for an area of maximum response was neither a success nor a failure. Time ran out before a definite evaluation of the technique could be made. However, the empirical results tend to substantiate the theoretical rationale on several points:

1) The design did move in the direction of increased response with a sequence of six experiments beyond the basic simplex resulting in a mean increase in the response of approximately one standard deviation.
2) The possibilities of including additional variables or rejecting ones that were having little effect was demonstrated by reducing the design from seven to four variables after the tenth experiment.
3) The ability of the experimental strategy to lead to more optimal results in the presence of experimental error was displayed when the strategy led to a rejection of the results of experiment 11 (where treatment effects were confounded with effects of prior instruction)

and a rerunning of that experimental combination.

4) The advantage of the simplex design over a factorial design in terms of reduced experimental effort was not shown. The advantage of the simplex procedure should not become apparent until one moves to the third or fourth level of some of the factors and the restricted length of the sequence did not permit this.

In retrospect, too much was attempted in this study. The attempt to utilize some of the ideas of response surface methodology in maximizing the response meant that the full potential for evaluating ideas from the concept learning literature in this setting was not realized. Similarly, the desire to use a task and response measure which would avoid the criticisms directed at other recent concept learning studies introduced delays which prevented a thorough test of the exploration ideas taken from the response surface literature. The simple expedient of using a measure of immediate learning rather than waiting twenty-four hours to test would have permitted the running of twice as many experiments in the same time period.

On the other hand, it should be noted that in neither area has this study been a failure; it has merely been

indecisive. More effort on the part of educational psychologists must be directed towards the study of concept learning with tasks which are more accurate analogs of what teachers think of as concepts. Furthermore, the identification of combinations of variables which, within specified restrictions, produce maximum learning would seem a necessary pre-requisite to the scientific planning of learning experiences. Educational researchers have not typically focussed on the problem of optimization of desired responses. However, it would seem that they should and it is suggested that a response surface conceptualization of many problems would be of value.

REFERENCES

Archer, E. S., Bourne, L. E., and Brown F.G. Concept identification as a function of irrelevant information and instructions. *J. Exptl. Psychol.*, 1955, *49*, 153-63.

Baker, Frank. Experimental design considerations associated with large-scale research projects. In Julian C. Stanley (Ed.), *Improving experimental design and statistical analysis*, Chicago: Rand McNally, 1967, 206-227.

Box, G.E.P. The exploration and exploitation of response surfaces. *Biometrics*, 1954, *10*, 16-60.

Box, G.E.P. Evolutionary operation: a method for increasing industrial productivity. *Applied Statistics*, 1957, *6*, 81-101.

Box, G.E.P. and Draper, N. R. A basis for the selection of a response surface design. *J. Amer. Statist. Assoc.*, 1959, *54*, 622-654.

Box, G.E.P. and Hunter, J. S. Multifactor experimental designs for exploring response surfaces. *Ann. Math. Statist.*, 1957, *28*, 195-241.

Box, G.E.P. and Hunter, J. S. Experimental designs for the exploration and exploitation of response surfaces. In V. Chew (Ed.), *Experimental designs in industry*. New York: Wiley, 1958.

Box, G.E.P. and Wilson, K. B. On the experimental attainment of optimum conditions. *J. Roy. Statist. Soc.*, *Series B*, 1951, *13*, 1-38.

Bradley, R. A. Determination of optimum operating conditions by experimental methods, Part I. *Indus. Qual. Control*, 1958, *15*, No. 4, 16-20.

Bruner, J. S., Goodnow, J. J., and Austin, G. A. *A Study of Thinking*. New York: Wiley, 1956.

Buck, R. Creighton. *Advanced Calculus*. New York: McGraw-Hill, 1956.

Davies, O. L. (Ed.) The design and analysis of industrial experiments. New York: Hafner, 1956.

Fisher, Sarah C. The process of generalizing abstraction; and its product, the general concept. Psychol. Monog., 1916, 21, No.2.

Glanzer, M., Huttenlocher, J. and Clark, W. H. Systematic operations in solving concept problems. Psychol. Monog., 1963, 77, No. 1.

Heidbreder, E. The attainment of concepts: I. Terminology and methodology. J. Gen. Psychol., 1946, 35, 173-189.

Hermanson, H. P. Maximization of potato yield under constraint. Agronomy Journal, 1965, 57, 210.

Hill, William J. and Hunter, William G. A review of response surface methodology: a literature survey. Technometrics, 1966, 8, 571-90.

Hotelling, Harold. Some improvements in weighting and other experimental techniques. Ann. Math. Stat., 1944, 15, 297.

Hovland, C. I. A "communication analysis" of concept learning. Psychol. Rev., 1952, 59, 461-472.

Hovland, C. I. and Weiss, W. I. Transmission of information concerning concepts through positive and negative instances. J. Exptl. Psychol., 1953, 45, 175-182.

Hull, Clark L. Quantitative aspects of the evolution of concepts. Psychol. Monog., 1920, 28, No. 1.

Hunt, E. B. and Hovland, C. I. Order of consideration of different types of concepts. J. Exptl. Psychol., 1960, 59, 220-225.

Hunter, J. S. Determination of optimum operating conditions by experimental methods, Part II-1. Indus. Qual. Control, 1958, 15, No. 6, 16-24.

Hunter, J. S. Determination of optimum operating conditions by experimental methods, Part II-2. Indus. Qual. Control, 1959a, 15, No. 7, 7-15.

Hunter, J. S. Determination of optimum operating conditions by experimental methods, Part II-3. Indus. Qual. Control, 1959b, 15, No. 8, 6-14.

Katos, S. L. and Yudin, L. Concept attainment and memory. J. Educ. Psychol., 1964, 55, 103-109.

Klausmeier, Herbert J., Davis, J. K., Ramsay, J. G., Fredrick, Wayne C., and Davies, M. H. Concept learning and problem solving: a bibliography. Technical Report No. 1 of Research and Development Center for Learning and Re-Education. Madison: University of Wisconsin, 1965.

Meyer, Donald L. Response surface designs with integer-valued factors. Unpublished doctoral dissertation. University of Minnesota, 1961.

Meyer, Donald L. Response surface methodology in education and psychology. J. Experimental Education, 1963, 31, 329-336.

Moore, T. V. The process of abstraction. University of California Publ. in Psychol., 1910, 1, 122.

Newman, James R. (Ed.) The world of mathematics. New York: Simon and Schuster, 1956.

Plackett, R. L. and Burman, J. P. The design of optimum multifactor experiments. Biometrika, 1946; 33,

Smith, H. and Rose, A. Subjective responses in process investigation. Indus. Eng. Chem., 1963, 55, 25.

Smoke, Kenneth L. An objective study of concept formation. Psychol. Monog., 1932, 42, No. 4.

Spendley, W., Hext, G. R. and Himsworth, F. R. Sequential application of simplex designs in optimization and evolutionary operation. Technometrics, 1962, 4, 441-462.

Weissert, F. C. and Cundiff, R. R. Compounding Diene/NR blends for truck tires. Rubber Age, 1963, 92, 881.

Welch, L. F., Adams, W. E., and Carmon, J. L. Yield response surfaces, isoquants and economic optima for coastal bermudagrass. Agronomy Journal, 1963, 55, 63.

Wittrock, M. C., Keislar, E. R., and Stern, C. Verbal cues in concept identification. J. Educ. Psych., 1964, 55, 195-200.

Wu, S. M. Tool life testing by response surface methodology, Part I. A.S.M.E. Transactions, Sec. B., J. Eng. Indus., 1964, 105-110.

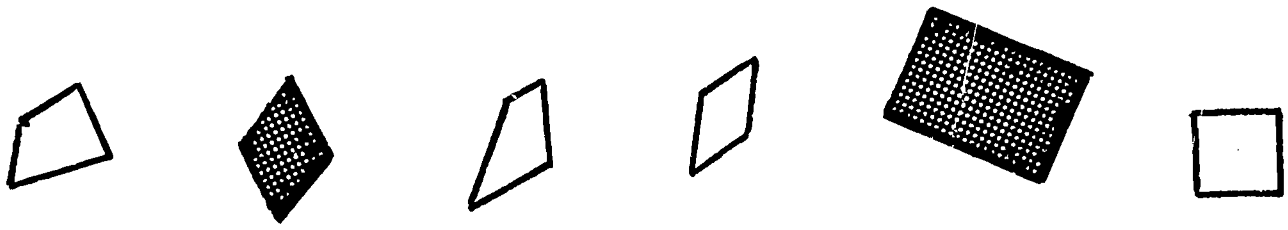
A P P E N D I C E S

APPENDIX A

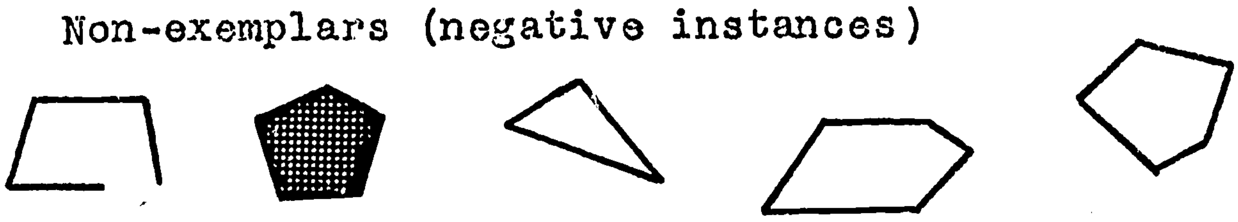
Figures used as exemplars and non-exemplars of the concepts quadrilateral, isosceles triangle, trapezoid and rectangle.

Quadrilateral

Exemplars (positive instances)

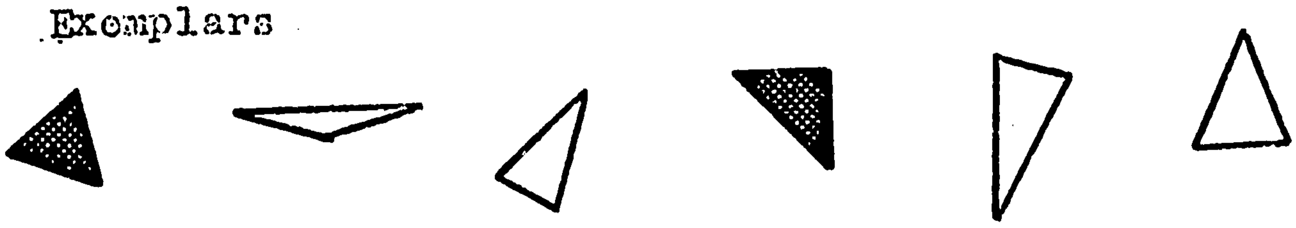


Non-exemplars (negative instances)



Isosceles Triangle

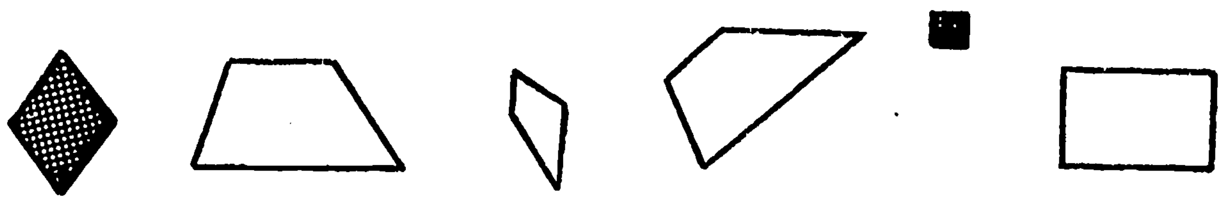
Exemplars



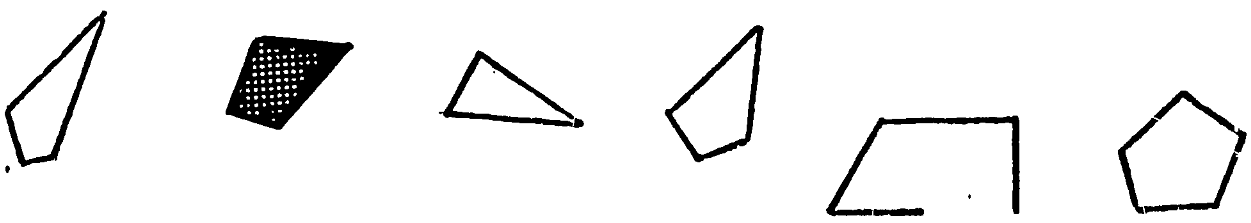
Non-exemplars

Trapezoid

Exemplars

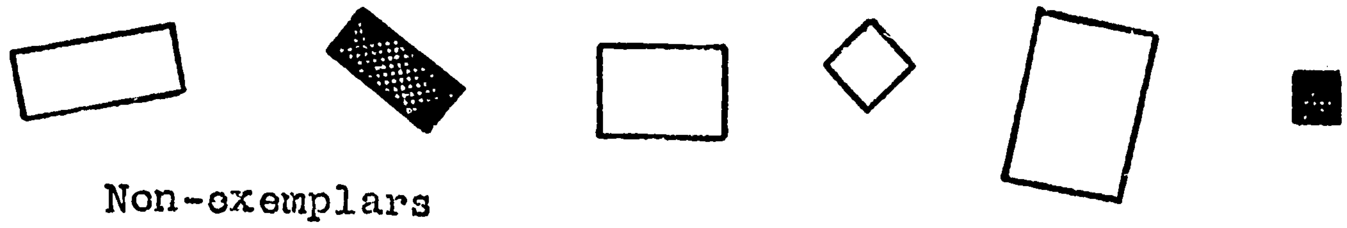


Non-exemplars

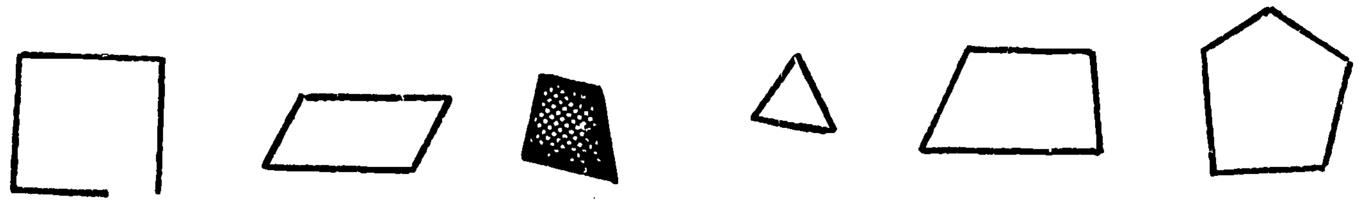


Rectangle

Exemplars



Non-exemplars



APPENDIX B

The following is the transcript of the instructions and warm-up tasks presented to all experimental groups. After an informal introduction by the experimenter, the following was presented via a tape recording.

You are taking part in an experiment to find out how well boys and girls your age can learn to identify certain kinds of geometric figures. We are also trying to determine what can be done to make it easier to learn these figures.

You may know something about some of the figures already. Even if you do we still want you to pay careful attention since some figures may have more than one name and we would like you to know all the names.

For example, this is a house. It is also a building. It is also a shelter.

(A line sketch of a house was projected.)

This, also, is a building and a shelter but you would not call it a house.

(The house was replaced by a sketch of a tent.)

This tent is a shelter but it is not a building nor is it a house.

Now you are going to see some short series of slides each showing a figure. You will be told whether

the figure is or is not an example of the kind of figure you are trying to learn. Your job will be to try to decide just what it is that gives a figure its name.

The first two series are practice ones about figures which all of you know. They will give you some idea of what to expect from the more difficult sets that follow.

(A series of five triangles was projected on the screen for eight seconds each with four seconds between each slide. Accompanying each slide was the verbal direction . . .)

This is a triangle.

(A pause of 15 seconds followed the last slide.)

That was an easy series. You should have been able to decide that a triangle is a figure or shape with three straight sides.

The next set of slides will be somewhat different. Pay careful attention to the first slide since it will be the only one to show you what the figure is. All the other sides will show you figures that do <u>not</u> fit the name.

(The subjects were then shown a sequence of five slides, the first of which was a square while the others were not. Each was accompanied by the verbal direction . . .)

This is (is not) a square.

Are you still certain you know what a square is? Could you put it in words?

(Pause for 15 seconds)

A square is a four-sided figure. All four sides are the same length. It also has four corners or angles all of which are alike.

We call this kind of corner or angle a "square corner" or "right angle."

(A slide with two lines intersecting in a right angle was projected.)

The remaining sets of slides will be more difficult since you probably will not know the words as you did with triangle and square. Remember that some figures may have more than one name. As I said before "all houses are buildings" but "not all buildings are houses." The same thing applies to some of the figures you will see.

Try to learn as much as you can from each set of slides. After each set you will not be told what the correct answer is, but you will be asked a question to see if you have understood what kind of figure fits the name.

(Pause)

In this first series of slides you will be trying to find out what a quadrilateral is.

(This was followed by the particular sequences of slides comprising the experimental treatments for the concept, quadrilateral.)

You should now have a good idea what it is that makes a quadrilateral. Could you draw one? Using your ruler try to draw a quadrilateral in the space after number 1 on the paper you were given.

(Pause 25 seconds.)

Now try to draw another one right next to it, but make this quadrilateral as different as you can from the first one and still have it be a quadrilateral.

(Pause 25 seconds.)

Now let us look at another set of slides. This time you will be trying to decide what a trapezoid is.

(This was followed by the sequence of slides on the concept, trapezoid.)

Could you now say what it is that makes a shape or figure a trapezoid? Try writing out in words what you think a trapezoid is. Write your definition after number 2.

(Pause 45 seconds.)

Let us try another set of slides. You will be trying to learn what an isosceles triangle is.

(This was followed by the sequence of slides on the concept, isosceles triangle.)

What makes an isosceles triangle a special kind of

triangle? Were you able to figure it out? Using your ruler draw an isosceles triangle after number 3 on your paper.

(Pause 30 seconds.)

Let's see if you can draw another one. Make this isosceles triangle as different as you can from the first one and still have it be an isosceles triangle.

(Pause 30 seconds.)

We have one more series of slides to show you today. This word may be more familiar than the last three but it will not be any easier to learn. You will be trying to decide what a rectangle is.

(This was followed by the sequence of slides on the concept, rectangle.)

Now you should know what a rectangle is. Write your definition of a rectangle after number 4.

(Pause 45 seconds.)

Let's see if you can spell all the words you learned today. Turn your papers over. After number 5 write the word, trapezoid---trapezoid. That is a hard one but try your best.

(Pause.)

After number 6 write the word quadrilateral---quadrilateral.

(Pause.)

After number 7 write the word rectangle---rectangle.

(Pause.)

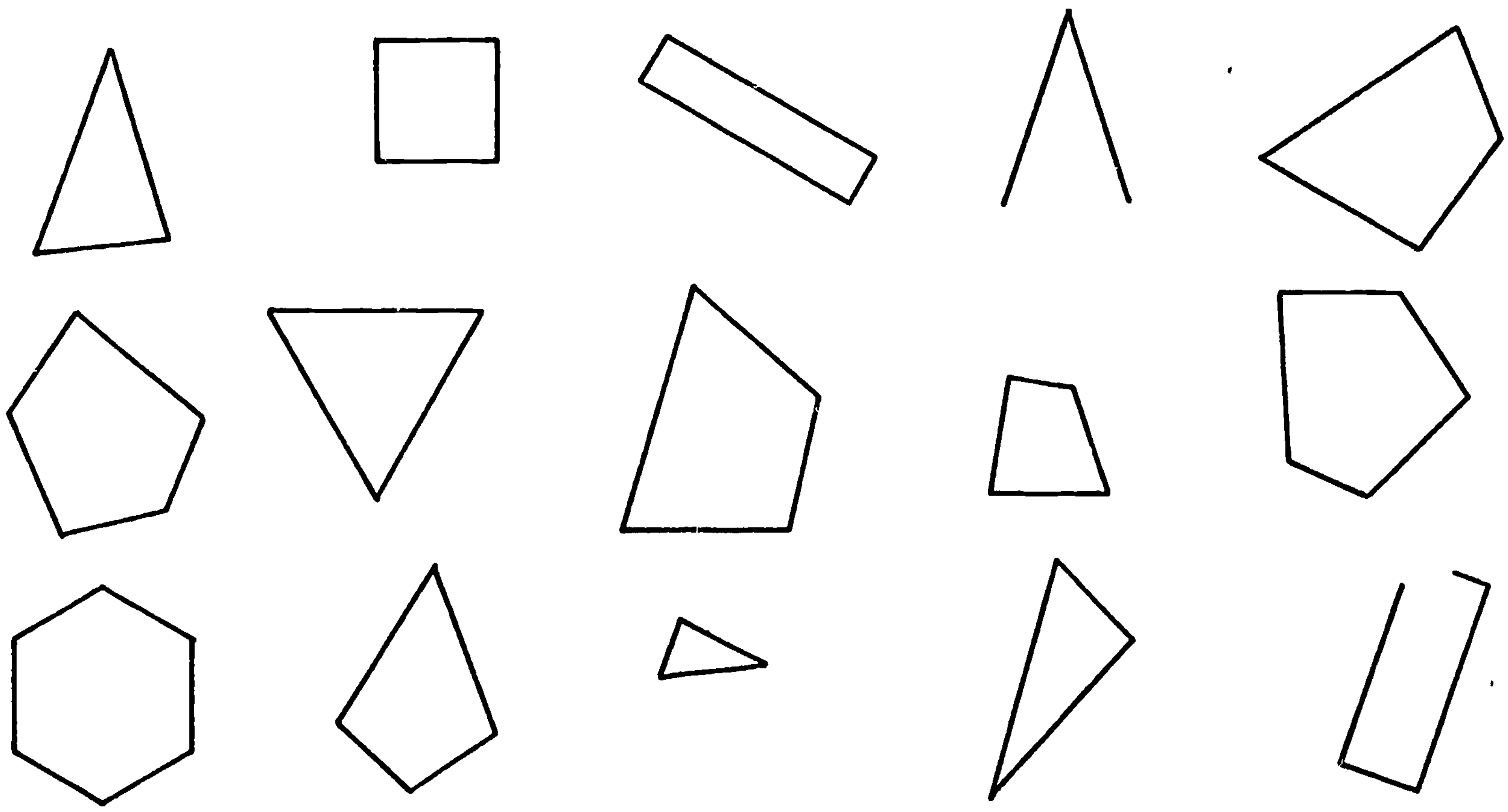
Now the last one is a little harder. Try writing the words isosceles triangle after number 8---isosceles triangle.

(At this point the tape recorder was turned off, the students were cautioned not to discuss the treatment until after the experimenter's return the next day, and sent back to their classrooms.)

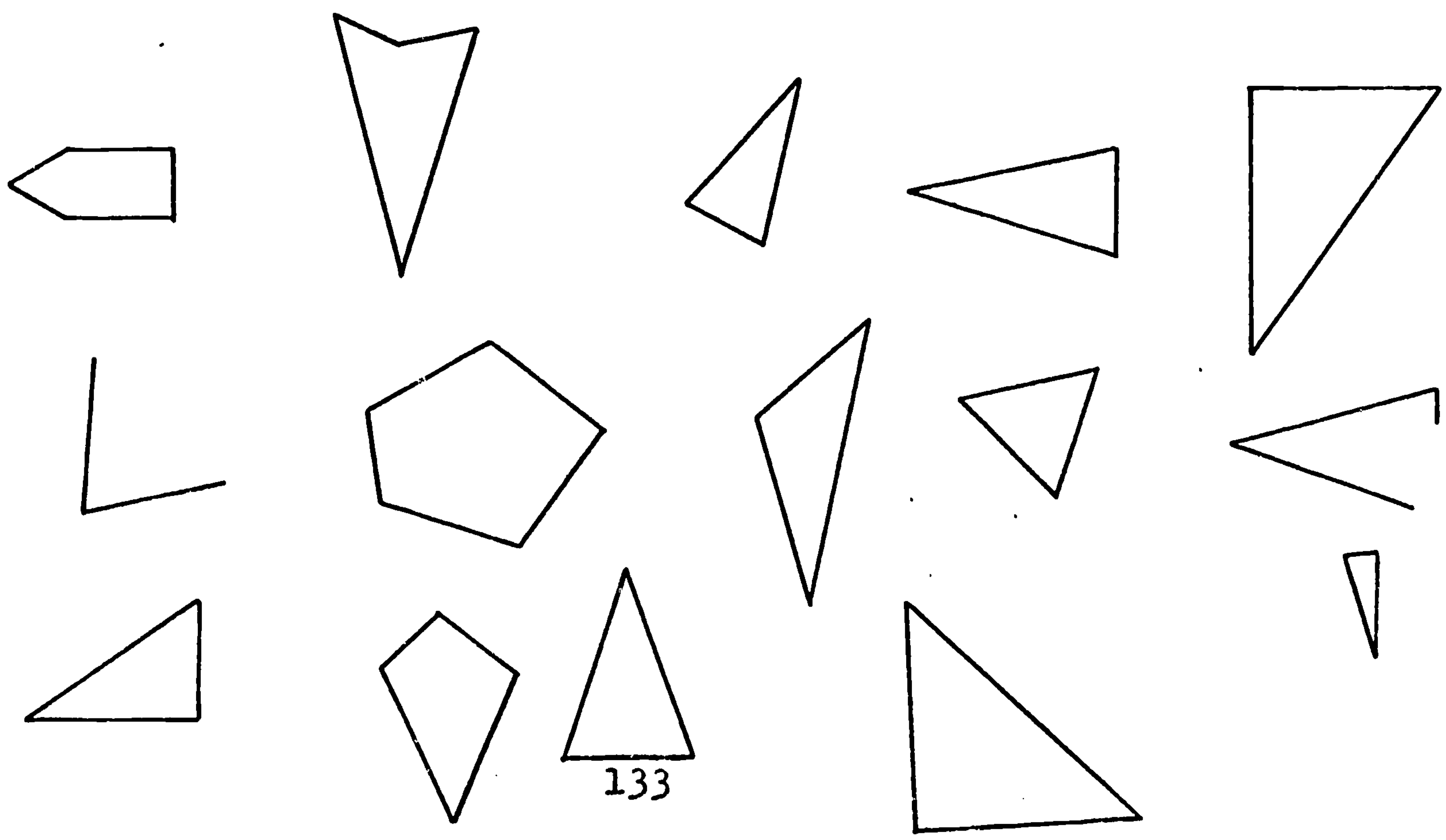
APPENDIX C

GEOMETRIC FIGURE IDENTIFICATION TEST

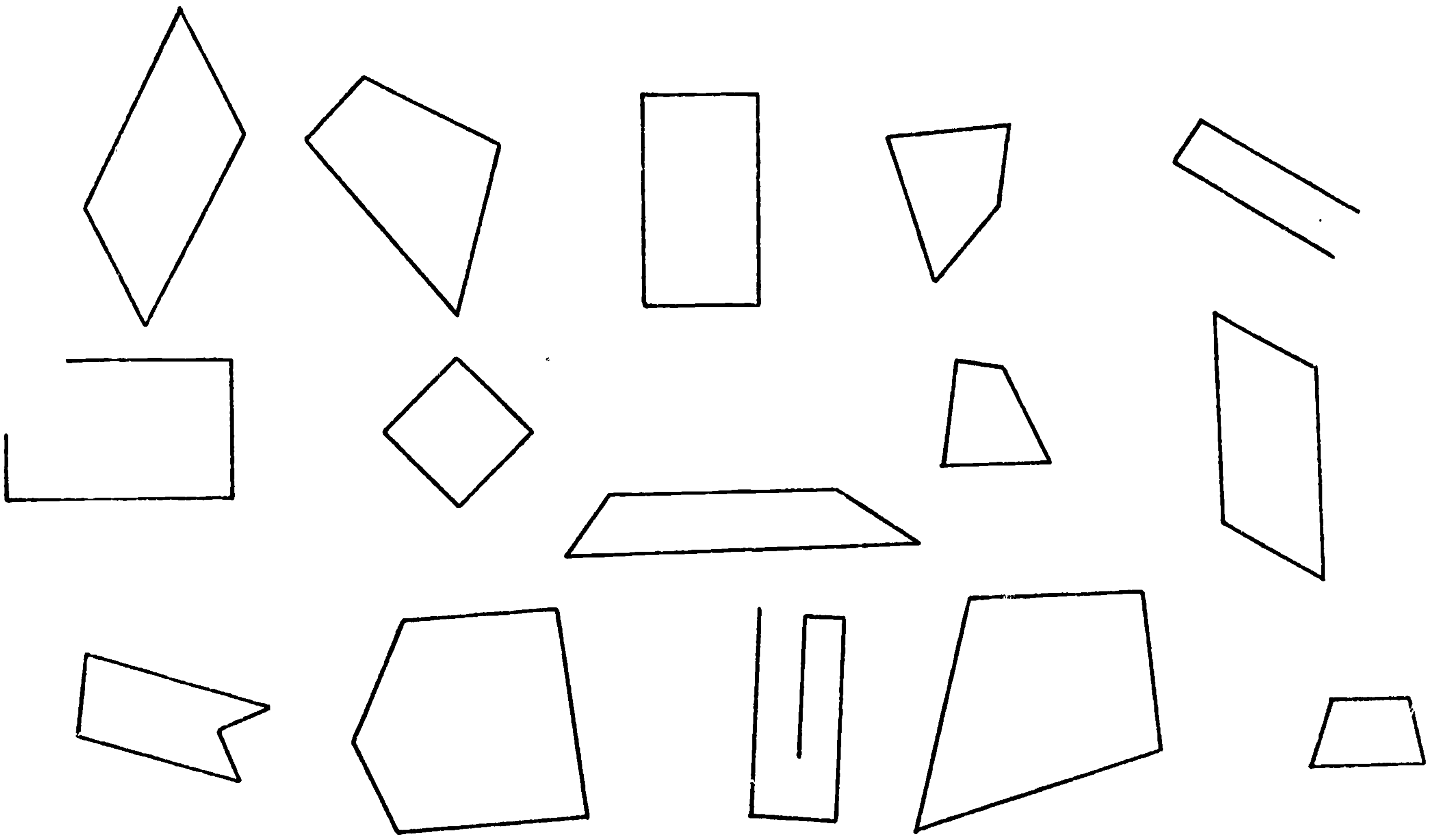
1. Circle each figure below which is a QUADRILATERAL. Do nothing to the other figures.



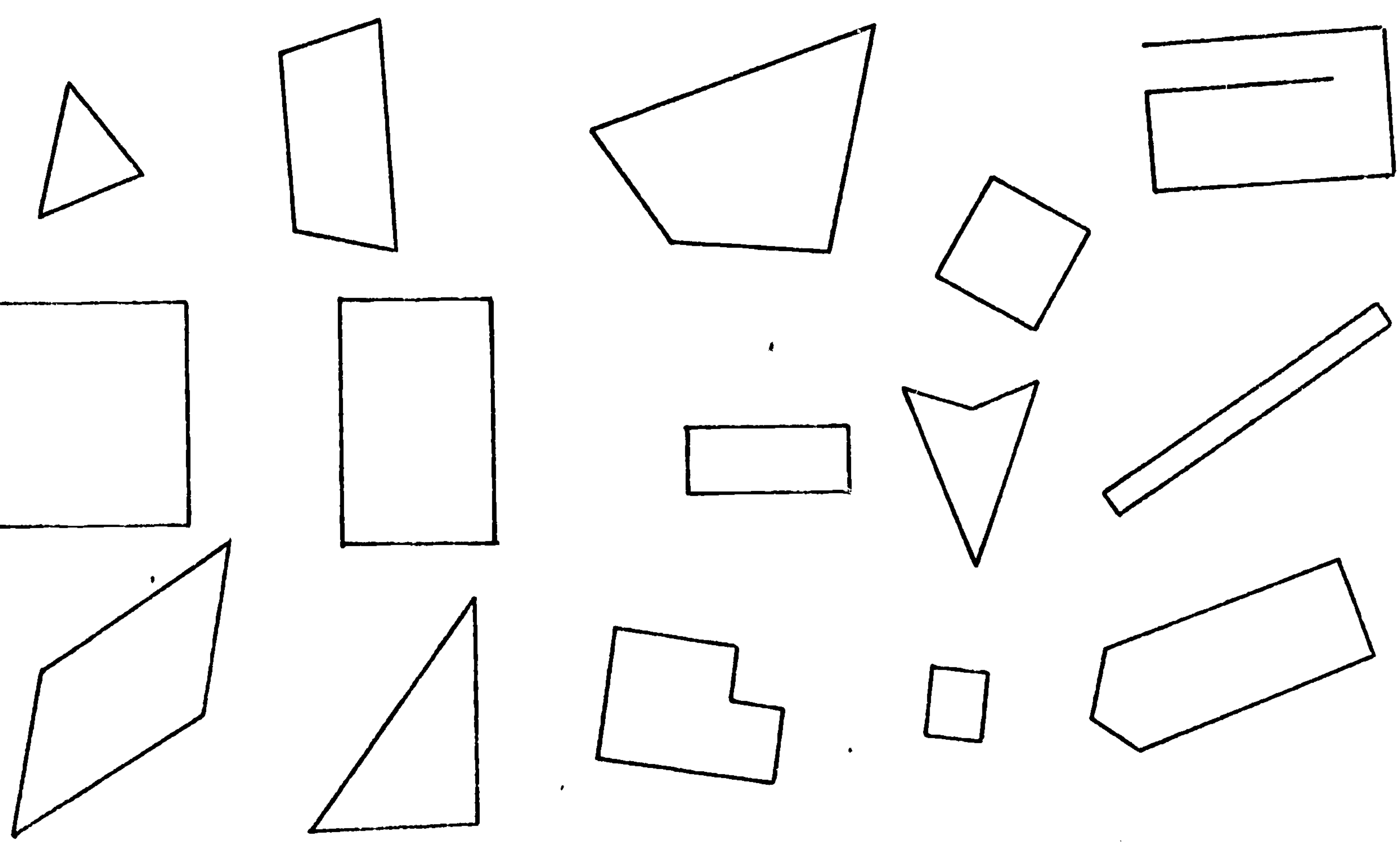
2. Circle each figure below which is an ISOSCELES TRIANGLE. Do nothing to the other figures.



3. Circle each figure below which is aTRAPEZOID. Do nothing to the other figures.



4. Circle each figure below which is a RECTANGLE. Do nothing to the other figures.



APPENDIX D

LETTER SOLICITING PARTICIPATION BY SCHOOL DISTRICTS

THE UNIVERSITY OF WISCONSIN
2218 University Avenue
Madison, Wisconsin 53705

Research and Development Center
For Learning and Re-Education March 17, 1965

Dean Lindley Stiles, Professor Herbert J. Klausmeier, and Mr. Chester W. Spangler of the Research and Development Center for Learning and Re-Education of the University of Wisconsin invite your cooperation in a research project designed to increase understanding of factors that affect children's learning of concepts. We propose to learn how various factors connected with presenting information to fifth- and sixth-grade children affect how well they learn and remember concepts in geometry through a single short audio-visual (slides and tape recording) presentation of plane geometry materials. We hope that this information will be useful in improving learning mathematics in the elementary school. Your cooperation would include the following:

1) Fill out the short information form which you find enclosed and return to the project director.

2) Furnish the project director with a list of your fifth-grade students prior to the day of the experiment. This will allow him to randomly assign students to experimental and control groups.

3) Allow the student's experimental team to work with the selected experimental group (approximately 20) for one hour on a prescheduled day. A room which is adequate both for the viewing of slides and writing will be needed for the experiment. The remainder of the fifth graders would be left with your teachers in another room.

4) Allow the experimental team to return the following day to administer to <u>all</u> fifth graders a short test covering the materials presented.

. This is all your cooperation will require. We have attempted to make the intrusion into the normal school program as short and unobtrusive as possible.

We should like to emphasize that we have absolutely no interest in making comparisons either between teachers or among schools. A summary report of the project will be sent to all schools when it becomes available.

Robert C. Remstad, a project assistant on the Center staff, will be in charge of the experiment. He will contact you within the next ten days to answer any questions you may have and to enlist your participation. We trust you will allow us to use some of your students for this purpose.

Yours sincerely,

Herbert J. Klausmeier, co-director
Research and Development Center
for Learning and Re-Education

APPENDIX E

The following Wisconsin elementary schools participated in this study by allowing the experimenter to use their fifth grade students as subjects for the sequence of experiments.

Black Earth--Mazomanie
- Black Earth Elementary
- Mazomanie Elementary

Blanchardville
- Blanchardville Elementary

Boscobel
- Boscobel Elementary

DeForest
- DeForest Elementary

Dodgeville
- Dodgeville Elementary

Elroy--Kendall--Wilton
- Elroy Elementary

Fennimore
- Fennimore Elementary

Hillsboro
- City Elementary

McFarland
- McFarland Elementary

Mauston
- LaCrosse Street Elementary

Mount Horeb
- Mount Horeb Elementary

New Glarus
- New Glarus Elementary

New Lisbon
- New Lisbon Elementary

Oakfield
- Oakfield Elementary

Stoughton
- Kegonsa Elementary

Sun Prairie
- Northside Elementary
- Southside Elementary

Verona
- Verona Elementary